ESTATÍSTICA II Ec/Fi

Econometria – Parte 1 (cap 2, 3 e 4 do Wooldridge)

O que é?

- Na origem: Análise quantitativa dos fenómenos económicos pelo desenvolvimento simultâneo da teoria e da observação recorrendo a procedimentos de inferência adequados.
- Hoje em dia mais virada para a aplicação de técnicas estatísticas à análise de dados económicos, financeiros, sociais, etc. com o objetivo de estimar relações entre uma determinada variável dependente e um conjunto de variáveis explicativas

- Exemplos:
    - Consumo= f(rendimento disponível)
    - Salário=f(escolaridade, experiência, idade, género)

Econometria

Finalidade da Econometria:

- Testar a validade de teorias
  - Será que o salário depende do nível de escolaridade, da experiência profissional, da idade ou do género?

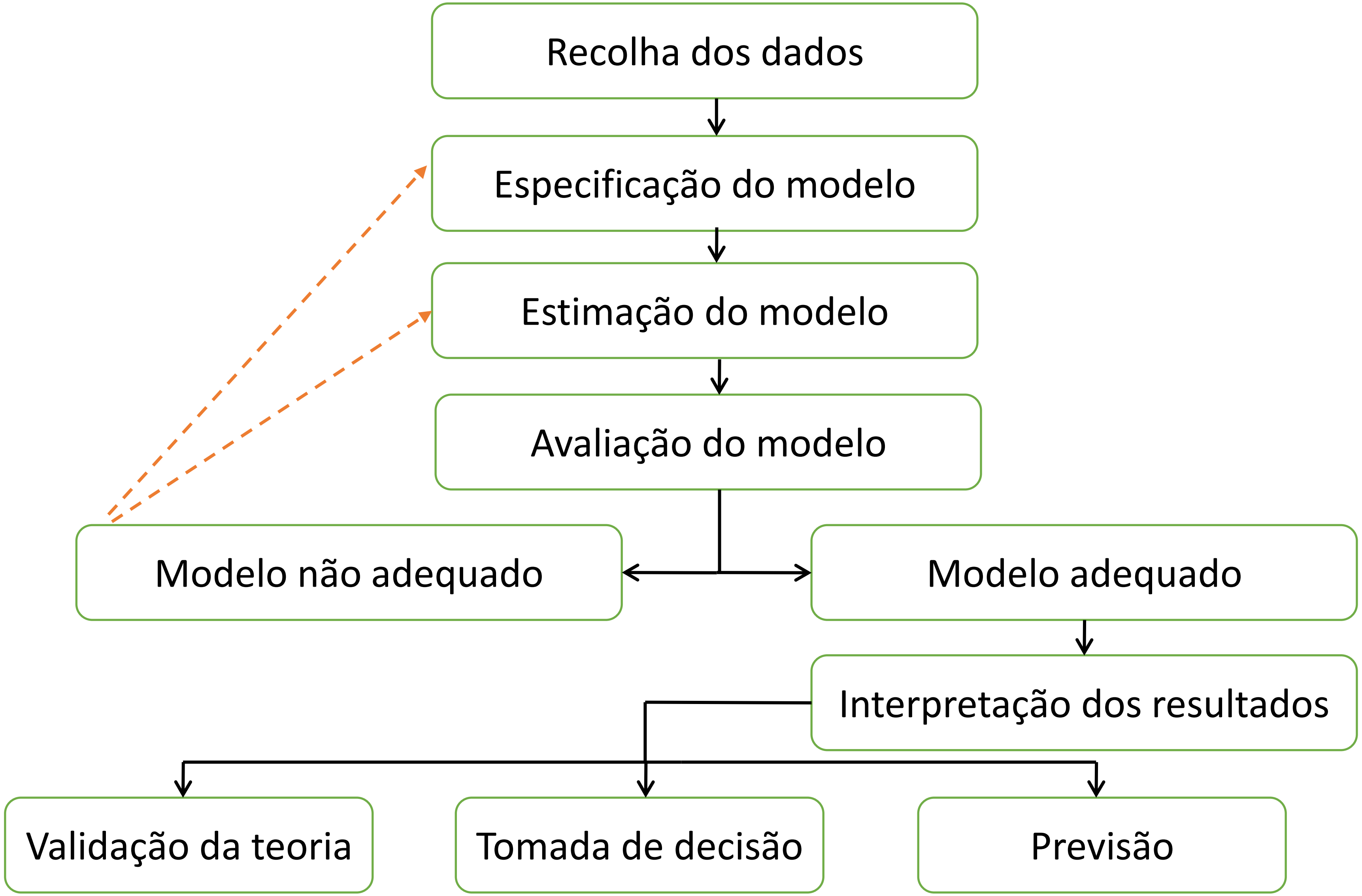
  Salário = f(escolaridade, experiência, idade, género)

- Efetuar previsões
  - Qual o salário esperado para um individuo com certas características?

- Fundamentar quantitativamente políticas
  - Aumentar o nível de escolaridade de certa população tem consequências na sua produtividade?

  Produtividade = f(escolaridade, experiência, idade, género)

Metodologia



Tipos de Dados

Dados seccionais (*cross-section*):

- N entidades (indivíduos, empresas, famílias, ….)
- 1 observação por entidade

Dados temporais (*time serie*):

- 1 entidade (empresa, país, …)
- T observações por entidade

Dados de painel (*panel data*):

- N entidades
- T observações por cada entidade

Outros tipos de dados como o “pooled data”

Dados Seccionais

- Exemplo: Salário = f(escolaridade, experiência, idade, género) a ser estimado com base numa amostra de N indivíduos (1 observação por indivíduo);
- O tempo está “fixo” – Todas as observações se referem ao “mesmo” período temporal;
- Neste quadro é possível por vezes considerar-se que a amostra corresponde a **observações independentes** com **idêntica variância** já que, como se irá ver, a média vai ser dada pelo modelo.
- Mas … nem sempre a situação é tão “simples”: Podem aparecer vários problemas quer ao nível do modelo (por exemplo a variância não ser a idêntica para todas as observações) quer dos dados (alguns casos não são observáveis ou são observáveis com erro)

Dados Temporais

- Exemplo: Modelo que relacione a Taxa de Juro a 3 meses dos títulos do tesouro de determinado país (TxJuro) com a inflação (inf) e o deficit orçamental (def), isso é TxJuro = f(inf, def). Dispõe-se de uma amostra cobrindo T anos;

- A entidade está fixa e observa-se o comportamento ao longo do tempo;

- As observações estão **ordenadas** ao contrário do que sucede com dados seccionais; A periocidade costuma ser constante (anos, trimestres, meses, semanas,…) embora possam existir "falhas" na série de valores.

- Em termos estatísticos, o primeiro problema que surge é que quase nunca se pode considerar que períodos de tempo consecutivos constituem observações independentes e portanto ter-se-á de recorrer a modelos mais complexos;

- A ser estudado na unidade curricular de Econometria

Causalidade e significância estatística

- O modelo procura identificar os **determinantes** do comportamento de uma variável de interesse: O que explica o salário de um trabalhador? Quais as características que determinam o preço de um imóvel?

- Na avaliação do modelo
  - verifica-se se as variáveis explicativas consideradas são **estatisticamente significativas** para explicar o comportamento da variável de interesse, na procura de uma **possível relação de causalidade**

  - **Causalidade e significância estatística -** . A causalidade vai para além da significância estatística já que envolve considerações teóricas

Análise *ceteris paribus*

- Num modelo envolvendo várias variáveis explicativas o foco situa-se, por vezes, nos efeitos de uma delas sobre a variável dependente.
    - No modelo Salário = f(escolaridade, experiência, idade, género) o nosso interesse pode estar no estudo da discriminação de género, isto é, no efeito específico da variável género no salário

- Obtém-se o **efeito parcial / marginal** de cada variável explicativa sobre a variável de interesse, sob a condição de ***ceteris paribus***, isto é, assumindo que tudo o resto se mantém constante.
    - Pode interessar apenas o sinal do efeito (se positivo / negativo existe uma relação direta / inversa com a variável dependente) ou também a sua intensidade

Enquadramento do modelo

- Modelo com **dados seccionais** (mesmo período temporal)

$$y = f(x_1, x_2, \dots, x_k)$$

- Os dados têm natureza não experimental (estão fora do controle de quem vai especificar o modelo) logo quer a variável y quer as variáveis explicativas $x_1, x_2, \dots, x_k$ vão ser consideradas aleatórias.

- Para aliviar a notação vai abandonar-se a convenção em torno das maiúsculas/minúsculas. As letras maiúsculas corresponderão agora a vetores ou matrizes.

O objetivo

- Modelo $\rightarrow y = f(x_1, x_2, \dots, x_k)$

  Salário=f(educação,experiência,idade,sexo)

- Das várias características da variável y, vai modelar-se o **valor esperado** de y **condicionado** por $x_1, x_2, \dots, x_k$. Relação estatística versus relação matemática

- Um modelo é sempre uma representação simplificada da realidade que permite sublinhar os determinantes mais importantes do comportamento da variável y.

- Assim, existem sempre variáveis explicativas (que se pensam menos relevantes) que não são incluídas no modelo.

Modelo de Regressão Linear Múltipla (MRLM)

- Objetivo genérico: explicar $E(y|\boldsymbol{x})$
- **Regressão linear**:$E(y|\boldsymbol{x})$ é função linear de um conjunto de parâmetros β, isto é

$$E(y|\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

- O modelo é linear nos parâmetros e não nas variáveis, isto é, pode acontecer que $x_2 = x_1^2$ ou $x_3 = \ln(x_2)$.
- Fazendo $u = y - E(y|\boldsymbol{x})$, pode escrever-se

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

Especificação do MRLM

- O modelo vai ser estimado com base numa amostra com N observações. Recorda-se que cada observação é dada por

$$(y_i, x_{i1}, x_{i2}, \dots, x_{ik}) \text{ para } i = 1,2, \dots, N$$

- Tem-se então

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i \qquad (i = 1, \cdots, N)$$

Apenas as variáveis y e $x_1, x_2, \cdots x_k$ são observáveis (*N* obs.)

u: termo de erro

β: parâmetros – serão estimados

k: nº variáveis explicativas

$k+1$: nº parâmetros (em modelos incluindo termo independente)

N: nº observações

Especificação do MRLM

Em termos matriciais vem

$$E(Y|\boldsymbol{X}) = \boldsymbol{X}\beta \quad \text{, ou} \quad Y = \boldsymbol{X}\beta + U$$

Com

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{N1} & \cdots & x_{Nk} \end{bmatrix}, \quad U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

Y e U são vetores com N linhas (nº observações)

$\boldsymbol{X}$ matriz $(N \times (k+1))$ - cada linha uma observação, 1ª coluna é o termo independente e depois cada coluna cada uma variável

β vetor com $k+1$ linhas (coeficientes da regressão)

Especificação do MRLM

Exemplo: Como interpretar u?

,

$$salário = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 idade + \beta_4 sexo + u$$

$U = Y - E(Y|\boldsymbol{X})$ diferença entre Y e o seu valor esperado concionado.

u também pode ser interpretado como o efeito no salário das variáveis explicativas que não estão explicitadas no modelo, fixadas essas, claro.

Estimação dos coeficientes

Modelo: $E(y|X) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$

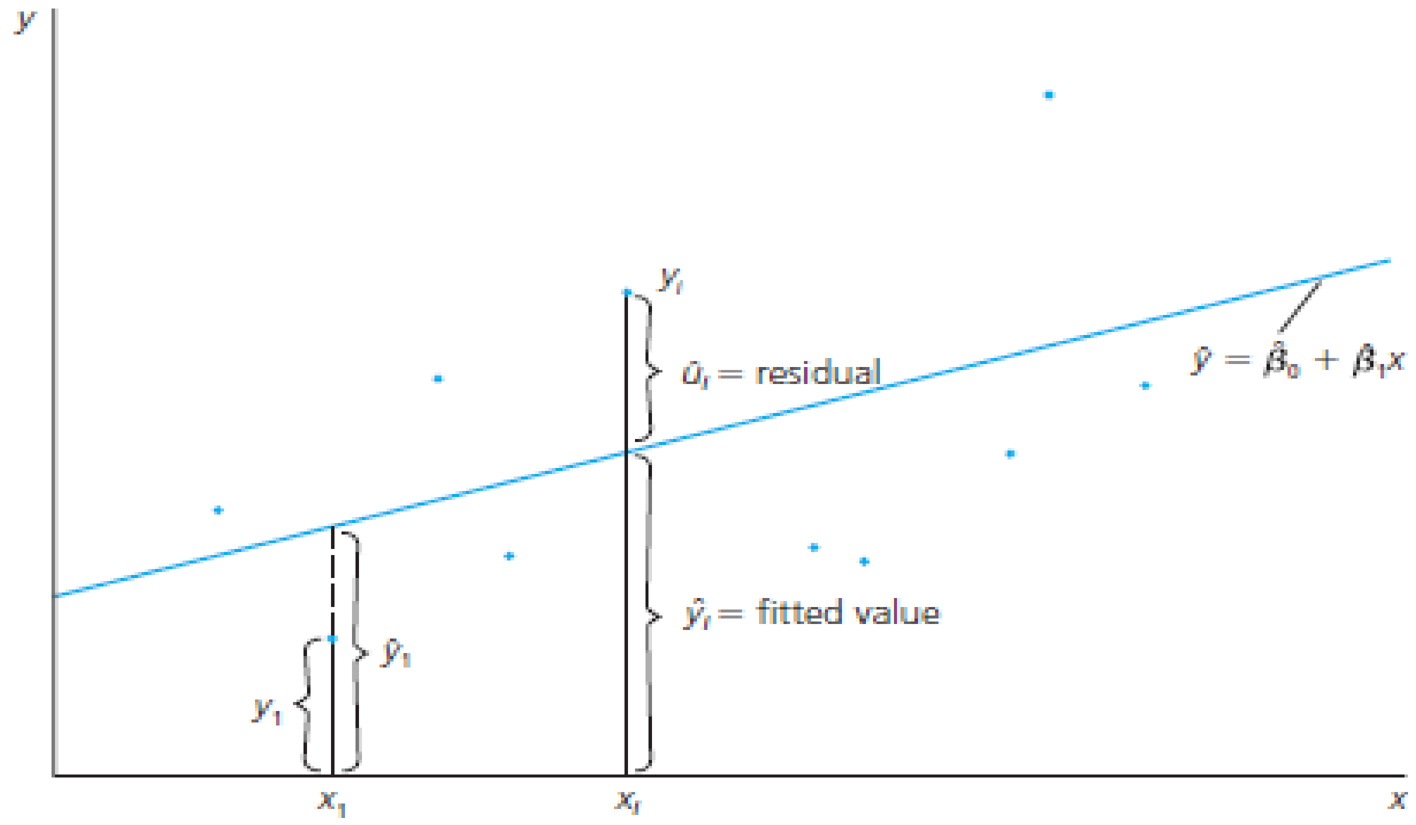
Amostra: $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i \quad (i = 1, \cdots, N)$

ou, na forma matricial, $Y = X\beta + U$

- Existindo $k + 1$ parâmetros desconhecidos a nossa 1ª preocupação será estimá-los
- Designe-se então por $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$ as estimativas obtidas a partir das quais se obtém:
    - Valores ajustados: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$
    - Resíduos: $\hat{u}_i = y_i - \hat{y}_i$
- Vamos escolher os valores $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$ que minimizam $\sum_{i=1}^{N} \hat{u}_i^2$

Valores ajustados e resíduos

Ilustração com um regressor (Wooldridge)



Variável residual versus resíduos

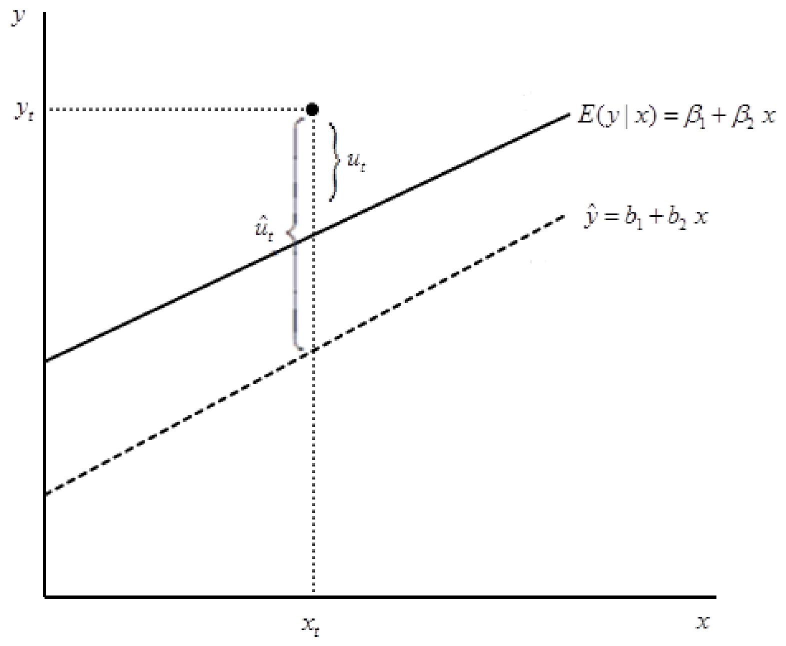


Figura tirada do capítulo 10 do livro de Estatística (Murteira *et al*)

Cuidado com a notação que é diferente da utilizada no Wooldridge

Estimação do Modelo:

- Método dos Mínimos Quadrados (OLS): $min \sum_{i=1}^{N} \hat{u}_i^2$
- *Como* $\sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} (y_i - \widehat{y_i})^2$

$$= \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ \dots - \hat{\beta}_k x_{ik}\right)^2$$

as condições de primeira ordem vêm (derivadas parciais em ordem a cada $\hat{\beta}$):

$$\begin{cases} \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ \dots - \hat{\beta}_k x_{ik}\right) = 0 \\ \sum_{i=1}^{N} x_{ij} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ \dots - \hat{\beta}_k x_{ik}\right) = 0 \qquad j = 1, \dots, k \end{cases}$$

OLS – Método dos Mínimos quadrados

- Em termos matriciais vem
  - Valores ajustados: $\hat{Y} = \boldsymbol{X}\hat{\beta}$
  - Resíduos: $\widehat{U} = Y - \hat{Y}$
- Método dos mínimos quadrados

$$\min \widehat{U}^T \widehat{U}$$

- *Condições de primeira ordem*

$$\boldsymbol{X}^T\left(Y - \boldsymbol{X}\hat{\beta}\right)=0$$

ou seja $\boldsymbol{X}^T Y = \boldsymbol{X}^T \boldsymbol{X}\hat{\beta}$

- Estimador MQ

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Estimação OLS

$$\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T Y$$

$$X^T X = \begin{bmatrix} N & \sum_{i=1}^{N} x_{i1} & \cdots & \sum_{i=1}^{N} x_{ik} \\ \sum_{i=1}^{N} x_{i1} & \sum_{i=1}^{N} x_{i1}^2 & \cdots & \sum_{i=1}^{N} x_{ik} x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} x_{ik} & \sum_{i=1}^{N} x_{ik} x_{i1} & \cdots & \sum_{i=1}^{N} x_{ik}^2 \end{bmatrix} \qquad X^T Y = \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_{i1} y_i \\ \vdots \\ \sum_{i=1}^{N} x_{ik} y_i \end{bmatrix}$$

Estimação OLS

Já que

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & x_{31} & \cdots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{n2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{1k} & x_{2k} & x_{3k} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ 1 & x_{31} & x_{32} & \cdots & x_{3k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

$$= \begin{bmatrix} N & \sum_{i=1}^{N} x_{i1} & \sum_{i=1}^{N} x_{i2} & \cdots & \sum_{i=1}^{N} x_{ik} \\ \sum_{i=1}^{N} x_{i1} & \sum_{i=1}^{N} x_{i1}^2 & \sum_{i=1}^{N} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{N} x_{i1}x_{ik} \\ \sum_{i=1}^{N} x_{i2} & \sum_{i=1}^{N} x_{i2}x_{i1} & \sum_{i=1}^{N} x_{i2}^2 & \cdots & \sum_{i=1}^{N} x_{i2}x_{ik} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{N} x_{ik} & \sum_{i=1}^{N} x_{ik}x_{i1} & \sum_{i=1}^{N} x_{ik}x_{i3} & \cdots & \sum_{i=1}^{N} x_{ik}^2 \end{bmatrix}$$

Estimação OLS

e que

$$X^TY = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & x_{31} & \cdots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{n2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{1k} & x_{2k} & x_{3k} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdots \\ y_n \end{bmatrix} . = \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_{i1}\, y_i \\ \sum_{i=1}^{N} x_{i2}\, y_i \\ \\ \sum_{i=1}^{N} x_{ik}\, y_i \end{bmatrix}$$

Caso particular 1 – Regressão linear simples

Modelo com **um** regressor: modelo linear simples

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

$$\hat{\beta} = \begin{bmatrix} \bar{y} - \hat{\beta}_1 \bar{x} \\ \dfrac{\sum_{i=1}^{N}(x_i - \bar{x})y_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \dfrac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \dfrac{s_{yx}}{s_x^2} \end{bmatrix}$$

A obtenção destes resultados é feita recorrendo, uma vez mais, às condições de 1ª ordem escritas agora para est situação particular:

$$\begin{cases} \sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\ \sum_{i=1}^{N} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^{N} y_i - N\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{N} x_i = 0 \\ \sum_{i=1}^{N} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{N} x_i - \hat{\beta}_1 \sum_{i=1}^{N} x_i^2 = 0 \end{cases}$$

Caso particular 1 – Regressão linear simples

Da 1ª equação vem $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

Utilzando este resultado na 2ª equação vem

$$\sum_{i=1}^{N} x_i y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^{N} x_i - \hat{\beta}_1 \sum_{i=1}^{N} x_i^2 = 0$$

$$\sum_{i=1}^{N} x_i y_i - \bar{y} \sum_{i=1}^{N} x_i + \hat{\beta}_1 \bar{x} \sum_{i=1}^{N} x_i - \hat{\beta}_1 \sum_{i=1}^{N} x_i^2 = 0$$

$$\sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y} + N\hat{\beta}_1 \bar{x}^2 - \hat{\beta}_1 \sum_{i=1}^{N} x_i^2 = 0$$

$$\sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y} - \hat{\beta}_1 \left(\sum_{i=1}^{N} x_i^2 - N\bar{x}^2\right) = 0$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} x_i y_i - N\bar{x}\bar{y}}{\left(\sum_{i=1}^{N} x_i^2 - N\bar{x}^2\right)} = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) y_i}{\sum_{i=1}^{N} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2} = \frac{s_{yx}}{s_x^2}$$

Caso particular 2 – Modelo só com *intercept*

Modelo **sem** regressores: só com termo independente (*intercept*)

$$y_i = \beta_0 + u_i$$
$$\hat{\beta}_0 = \bar{y}$$

Já que $\sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} \left(y_i - \widehat{\beta}_0\right)^2$

1ª derivada = 0: $-2\sum_{i=1}^{N}(y_i - \hat{\beta}_0) = 0 \Leftrightarrow -2\sum_{i=1}^{N} y_{i.} + 2N\hat{\beta}_0 = 0$

$$\Leftrightarrow \hat{\beta}_0 = \bar{y}$$

Caso particular 3 – Regressão sem *intercept*

Modelo **sem termo independente** ($\beta_0 = 0$): regressão na origem

$$y_i = \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i$$

Recorrer à expressão geral para obter $\hat{\beta}_j$

Ter presente que este modelo **não verifica** $\sum_{i=1}^{N} \hat{u}_i = 0$, o que vai ter pesadas consequências nas propriedades dos estimadores.

Desde já se pode adiantar que não incluir termo independente quando se deveria é **muito mais grave** do que incluir quando não se deveria. Assim, só se utiliza este modelo em situações muito particlares.

Estimação OLS - Exemplo

Exemplo: Considere o modelo

$$preço = \beta_0 + \beta_1 area + \beta_2 quartos + u$$

onde o **preço de uma casa**, em milhares de dólares, depende da **área** (m^2) e do **número de quartos** (dados adaptados do ficheiro HPRICE1 do Wooldridge – Os pés quadrados foram convertidos em m2 – e disponíveis no FENIX).

Wooldridge Source: Collected from the real estate pages of the Boston Globe during 1990. These are homes that were sold in the Boston, MA area.

Estimação OLS - Exemplo

Modelo estimado

$$\widehat{preço} = -19.286 + 1.384area + 15.121quartos$$

Output EXCEL

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.794906945					
R Square	0.63187705					
Adjusted R Square	0.623215334					
Standard Error	63.04837628					
Observations	88					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	2	579971.1994	289985.5997	72.95055817	3.58672E-19	
Residual	85	337883.3088	3975.097751			
Total	87	917854.5083				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-19.28550028	31.0475285	-0.621160563	0.536156232	-81.0163048	42.44530424
area(m2)	1.38360615	0.148943494	9.289470184	1.40049E-14	1.08746658	1.67974572
quartos	15.12133684	9.488597692	1.593632413	0.114730383	-3.744537442	33.98721112

Estimação OLS - Exemplo

Em termos matriciais, viria para este exemplo

Dados (ver ficheiro)

Obs	Preço	área	quartos
1	300.0	226	4
2	370.0	193	3
3	191.0	128	3
….			
86	202.5	146	3
87	219.0	110	2
88	242.0	165	4

Matrizes

$$\boldsymbol{X} = \begin{bmatrix} 1 & 226 & 4 \\ \dots & \dots & \dots \\ 1 & 165 & 4 \end{bmatrix}$$ **(88 linhas por 3 colunas)**

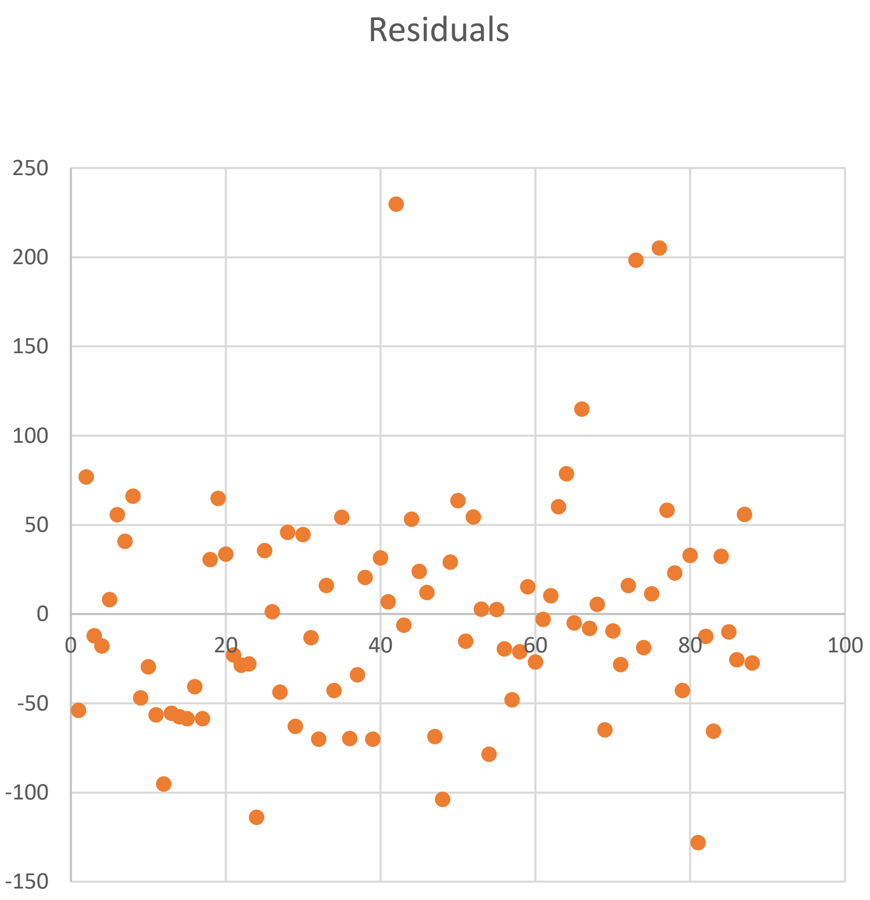
$$Y = \begin{bmatrix} 300 \\ \dots \\ 242 \end{bmatrix}$$ **(88 linhas)**

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} 88 & 16465 & 314 \\ 16465 & 3330603 & 60838 \\ 314 & 60838 & 1182 \end{bmatrix}$$ **mat simétrica**

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{bmatrix} 0.2424969362 & -0.0003691329 & -0.0454202476 \\ -0.0003691329 & 5.580785e-06 & -0.0001891845 \\ -0.0454202476 & -0.0001891845 & 0.0226493766 \end{bmatrix}$$

$$\boldsymbol{X}^T Y = \begin{bmatrix} 25832.05 \\ 5210658.92 \\ 95993.60 \end{bmatrix}$$ **vetor de dimensão 3**

Estimação OLS – Exemplo



A ordem da observação na amostra (i) está no eixo das abscissas enquanto os resíduos($\hat{u}_i$) estão no eixo das ordenadas.

$$\hat{u}_i = y_i - \hat{y}_i$$

Como para a 1ª obs, preço=300, area=226 e quartos=4 vem

$$\widehat{preço}_1 = -19.286 + 1.384 \times 226 + 15.121 \times 4 = 353.894$$

$\hat{u}_1$=300-353.894=-53.894

e, de forma semelhante,

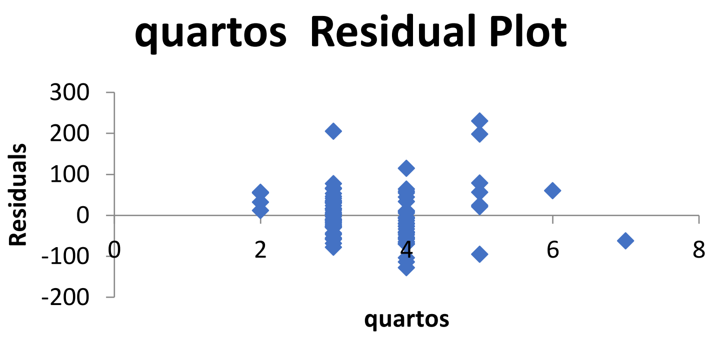
$\hat{u}_2$ = 370 - 293.114 = 76.884

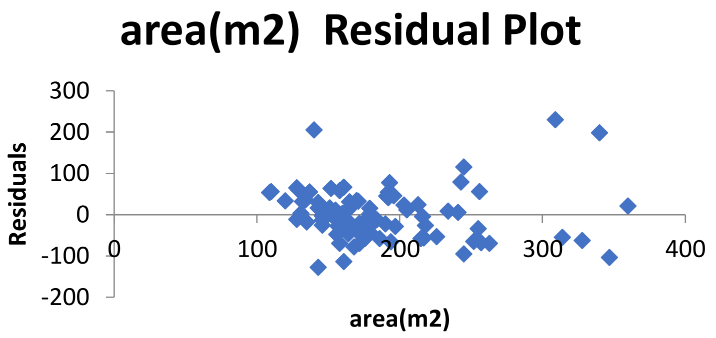
….

$\hat{u}_{88}$ =242 – 269.495 = –27.495

Estimação OLS – Exemplo

Como a ordem das observações é arbitrária num modelo seccional, torna-se muitas vezes mais interessante olhar para os resíduos em função de cada variável explicativa ou do valor da varíavel endógena.





Propriedades dos resíduos OLS

Resíduos: $\hat{u}_i = y_i - \hat{y}_i$

Propriedades:

- $\sum_{i=1}^{N} \hat{u}_i = 0$ (modelos com termo independente)
- $\sum_{i=1}^{N} \hat{u}_i x_{ij} = 0$, j=1,….k
- $\sum_{i=1}^{N} \hat{u}_i \hat{y}_i = 0$
- $\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} \hat{y}_i^2 + \sum_{i=1}^{N} \hat{u}_i^2$

Para demonstar estas propriedades é necessário recordar as condições de 1ª ordem que se obtiveram anteriormente:

$$\begin{cases} \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ \dots - \hat{\beta}_k x_{ik}\right) = 0 \\ \sum_{i=1}^{N} x_{ij} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ \dots - \hat{\beta}_k x_{ik}\right) = 0 \qquad j = 1, \dots, k \end{cases}$$

Propriedades dos resíduos OLS

$$\begin{cases} \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ \dots - \hat{\beta}_k x_{ik}\right) = 0 \\ \sum_{i=1}^{N} x_{ij}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ \dots - \hat{\beta}_k x_{ik}\right) = 0 \qquad j = 1, \dots, k \end{cases}$$

- **A propriedade 1** deriva diretamente da 1ª equação

$$\begin{aligned} \sum_{i=1}^{N} \hat{u}_i &= \sum_{i=1}^{N} (y_i - \hat{y}_i) \\ &= \sum_{i=1}^{N} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ \dots - \hat{\beta}_k x_{ik}\right) \\ &= 0 \ \ (\text{1ª eq. derivada parcial de } \hat{U}^T \hat{U} \text{ em ordem a } \hat{\beta}_0) \end{aligned}$$

- **A propriedade 2** deriva da "segunda" equação

$$\begin{aligned} \sum_{i=1}^{N} \hat{u}_i \, x_{ij} &= \sum_{i=1}^{N} x_{ij} (y_i - \hat{y}_i) \\ &= \sum_{i=1}^{N} x_{ij} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ \dots - \hat{\beta}_k x_{ik}\right) \\ &= 0 \ \ (\text{2ª eq. derivada parcial de } \hat{U}^T \hat{U} \text{ em ordem a } \hat{\beta}_j) \end{aligned}$$

Propriedades dos resíduos OLS

Propriedade 3 - $\sum_{i=1}^{N} \hat{u}_i \hat{y}_i = 0$

$$\sum_{i=1}^{N} \hat{u}_i \hat{y}_i = \sum_{i=1}^{N} \hat{u}_i \left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ \dots + \hat{\beta}_k x_{ik}\right)$$

$$= \hat{\beta}_0 \sum_{i=1}^{N} \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^{N} \hat{u}_i x_{i1} + \cdots + \hat{\beta}_k \sum_{i=1}^{N} \hat{u}_i x_{ik}$$
$$= 0$$

aplicando

propriedade 1 ($\sum_{i=1}^{N} \hat{u}_i$=0) e

propriedade 2 ($\sum_{i=1}^{N} \hat{u}_i x_{ij} = 0$ para $j = 1{,}2, \dots, k$)

Propriedades dos resíduos OLS

Propriedade 4 - $\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} \hat{y}_i^2 + \sum_{i=1}^{N} \hat{u}_i^2$

$$\begin{aligned}\sum_{i=1}^{N} y_i^2 &= \sum_{i=1}^{N} (\hat{y}_i + \hat{u}_i)^2 = \sum_{i=1}^{N} \hat{y}_i^2 + \sum_{i=1}^{N} \hat{u}_i^2 + 2\sum_{i=1}^{N} \hat{y}_i \hat{u}_i \\ &= \sum_{i=1}^{N} \hat{y}_i^2 + \sum_{i=1}^{N} \hat{u}_i^2\end{aligned}$$

já que , pela propriedade 3, $\sum_{i=1}^{N} \hat{y}_i \hat{u}_i = 0$

Qualidade do Ajustamento

Modelo: $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$

Modelo estimado: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1\, x_1 + \cdots + \hat{\beta}_k\, x_k$

Para fazer uma primeira avaliação à qualidade do ajustamento, pode comparar-se $\hat{y}_i$ com y_i e calcular o coeficiente de correlação

$$r_{y\hat{y}} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})\left(\hat{y}_i - \bar{\hat{y}}\right)}{\left(\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)\left(\sum_{i=1}^{n}\left(\hat{y}_i - \bar{\hat{y}}\right)^2\right)\right)^{1/2}} = \frac{s_{y\hat{y}}}{s_y\, s_{\hat{y}}}$$

Como o sinal da correlação não é informativo (será sempre positiva uma vez que $\hat{y}$ constitui a nossa melhor aproximação a y) utiliza-se o seu quadrado, $r_{y\hat{y}}^2$ (varia entre 0 e 1, quanto mais próximo de 1, melhor o ajustamento).

Decomposição da variação total e R^2

A abordagem mais habitual consiste em avaliar qual a proporção da variação da variável y que é "explicada" pelo modelo, isto é, coberta pela variação de $\hat{y}$.

A ideia é decompor a variação total (SST) em termos da variação explicada (SSE) e da variação residual (SSR)

Variação total (SST) = variação explicada (SSE)+ variação residual (SSR)

e ver o peso da variação explicada na variação total, definindo-se o coeficiente de determinação $\boldsymbol{R^2}$ como

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

Antes de comentar esta medida, veja-se o significado das diferentes componentes envolvidas.

Decomposição da variação total e R^2

Variação total (SST) = variação explicada (SSE)+ variação residual (SSR)

$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$

$SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$

$y_i = \hat{y}_i + \hat{u}_i$ logo $\sum y_i = \sum \hat{y}_i + \sum \hat{u}_i = \sum \hat{y}_i$ (termo ind) e portanto $\bar{y} = \bar{\hat{y}}$

$SSR = \sum_{i=1}^{n}(\hat{u}_i - \bar{\hat{u}})^2 = \sum_{i=1}^{n} \hat{u}_i^2$

como o modelo tem termo independente, $\sum \hat{u}_i = 0$ e portanto $\bar{\hat{u}} = 0$.

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{u}_i + (\hat{y}_i - \bar{y}))^2$$
$$= \sum \hat{u}_i^2 + \sum(\hat{y}_i - \bar{y})^2 + 2\sum \hat{u}_i(\hat{y}_i - \bar{y}) = SSR + SSE + 2\sum \hat{u}_i\hat{y}_i - 2\bar{y}\sum \hat{u}_i$$
$$= SSE + SSR$$

Recordam-se as **propriedades dos resíduos** (modelos com termo independente)

$\sum_{i=1}^{n} \hat{u}_i = 0$

$\sum_{i=1}^{n} \hat{u}_i\hat{y}_i = 0$

$\sum_{i=1}^{n} \hat{u}_i x_{ij} = 0$, j=1,….k

$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} \hat{y}_i^2 + \sum_{i=1}^{n} \hat{u}_i^2$

Quadro ANOVA

- Quadro de análise de variância: decomposição da variação total de Y

	SS	DF	MS
Explicados	$\sum(\hat{y}_i - \bar{y})^2$	k	MSE
Residual	$\sum \hat{u}_i{}^2$	$n - k - 1$	$MSR = \hat{\sigma}^2$
Total	$\sum(y_i - \bar{y})^2$	$n - 1$	s_y^2

- Coeficiente de determinação:

$$R^2 = \frac{SSE}{SST}, \qquad 0 \leq R^2 \leq 1$$

Mede a proporção da variação amostral de y explicada pela regressão

R^2 - Interpretação

Alguns comentários:

- Recorrendo às propriedades dos resíduos, mostra-se que no MRLM (com termo independente) $r_{y\hat{y}}^2 = R^2$.

$$\bar{\hat{y}} = \frac{\sum \hat{y}_i}{n} = \frac{\sum (y_i - \hat{u}_i)}{n} = \frac{\sum y_i}{n} - \frac{\sum \hat{u}_i}{n} = \bar{y}$$ Propriedade 1 dos resíduos OLS

$$\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) = \sum (\hat{y}_i - \bar{y} + \hat{u}_i)(\hat{y}_i - \bar{y})$$
$$= \sum \left((\hat{y}_i - \bar{y})^2 + \hat{u}_i(\hat{y}_i - \bar{y}) \right)$$
$$= \sum (\hat{y}_i - \bar{y})^2 + \sum \hat{u}_i \hat{y}_i - \bar{y} \sum \hat{u}_i$$
$$= \sum (\hat{y}_i - \bar{y})^2$$ Propriedades 1 e 3 dos resíduos OLS

$$r_{y\hat{y}}^2 = \frac{\left(\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) \right)^2}{\left(\sum_{i=1}^n (y_i - \bar{y})^2 \right) \left(\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \right)} = \frac{\left(\sum (\hat{y}_i - \bar{y})^2 \right)^2}{\left(\sum_{i=1}^n (y_i - \bar{y})^2 \right) \left(\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \right)}$$
$$= \frac{SSE}{SST} = R^2$$

R^2 - Interpretação

Alguns comentários:

- De um modo geral $SSE = R^2\, SST$ ou SSR $= (1 - R^2)\, SST$

$$SSE = R^2\, SST \Leftrightarrow SST - SSR = R^2\, SST \Leftrightarrow SST - R^2\, SST = SSR$$

- $R^2 = 1$ significa que $SSR = \sum_{i=1}^{n} \hat{u}_i^2 = 0$, isto é, que o ajustamento é perfeito $\hat{y}_i = y_i$.
- $R^2 = 0$ significa que $SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = 0$, isto é, que $\hat{y}_i = \bar{y}$ ($\hat{y}_i$ é constante) e que o modelo nada adianta.
- A interpretação do valor do R^2 levanta ainda 2 problemas:
  - Consoante o problema que se esteja a modelar, o R^2 assume valores bem diferentes: um valor de 0.3 tanto pode corresponder a um bom valor como a um mau modelo (depende do que se esteja a modelar)
  - Quando se acrescentam variáveis explicativas ao modelo, mostra-se que o R^2 não pode decrescer! Assim, o R^2 tende a valorizar modelos com excesso de variáveis explicativas.

R^2 ajustado

- A solução mais comum para corrigir este último problema consiste em recorrer ao coeficiente de determinação corrigido (Adusted R square ou $\bar{R}^2$) que “penaliza” o modelo pela introdução de novas variáveis explicativas.

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$$

O inconveniente desta nova medida é não ter uma interpretação tão intuitiva quanto o R^2.

Também se mostra que

$$\bar{R}^2 = 1 - (1 - R^2)\frac{(n-1)}{(n-k-1)}$$

R^2 - Comparação de modelos

- Formas alternativas do coeficiente de determinação e **comparação de modelos**:

	$R^2 = \frac{SSE}{SST}$	$r_{y\hat{y}}^2 = cor(Y, \hat{Y})^2$	$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k-1}$
Os modelos têm todos termo independente	V		V
Os modelos têm todos a mesma variável dependente (y)	V		V
O número de variáveis explicativas (k) é o mesmo	V	V	

Um R^2 elevado (baixo) não significa que o modelo descreva (não descreva) convenientemente os dados: será necessário testar a validade do modelo

Output Excel

Modelo estimado:

$$\widehat{preço} = -19.286 + 1.3836\ area + 15.121\ quartos$$

Identificar: $r_{y\hat{y}} = 0.7949$ $R^2 = 0.6319$ $\bar{R}^2 = 0.6232$ $n = 88$

$SST = 917854.5$ $SSE = 579971.2$ $SSR = \sum \hat{u}_i^2 = 337883.3$

$$\hat{\beta} = \begin{bmatrix} -19.286 \\ 1.3836 \\ 15.121 \end{bmatrix}$$

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.794906945					
R Square	0.63187705					
Adjusted R Square	0.623215334					
Standard Error	63.04837628					
Observations	88					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	2	579971.1994	289985.5997	72.95055817	3.58672E-19	
Residual	85	337883.3088	3975.097751			
Total	87	917854.5083				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-19.28550028	31.0475285	-0.621160563	0.536156232	-81.0163048	42.44530424
area(m2)	1.38360615	0.148943494	9.289470184	1.40049E-14	1.08746658	1.67974572
quartos	15.12133684	9.488597692	1.593632413	0.114730383	-3.744537442	33.98721112

Output Eviews

Modelo estimado:

$$\widehat{preço} = -19.286 + 1.3836\, area + 15.121\, quartos$$

Identificar: $R^2 = 0.6319$ $\bar{R}^2 = 0.6232$ $n = 88$ $SSR = \sum \hat{u}_i^2 = 337883.3$

$$\hat{\beta} = \begin{bmatrix} -19.286 \\ 1.3836 \\ 15.121 \end{bmatrix}$$

Dependent Variable: PRECO
Method: Least Squares
Date: 03/24/20 Time: 17:00
Sample: 1 88
Included observations: 88

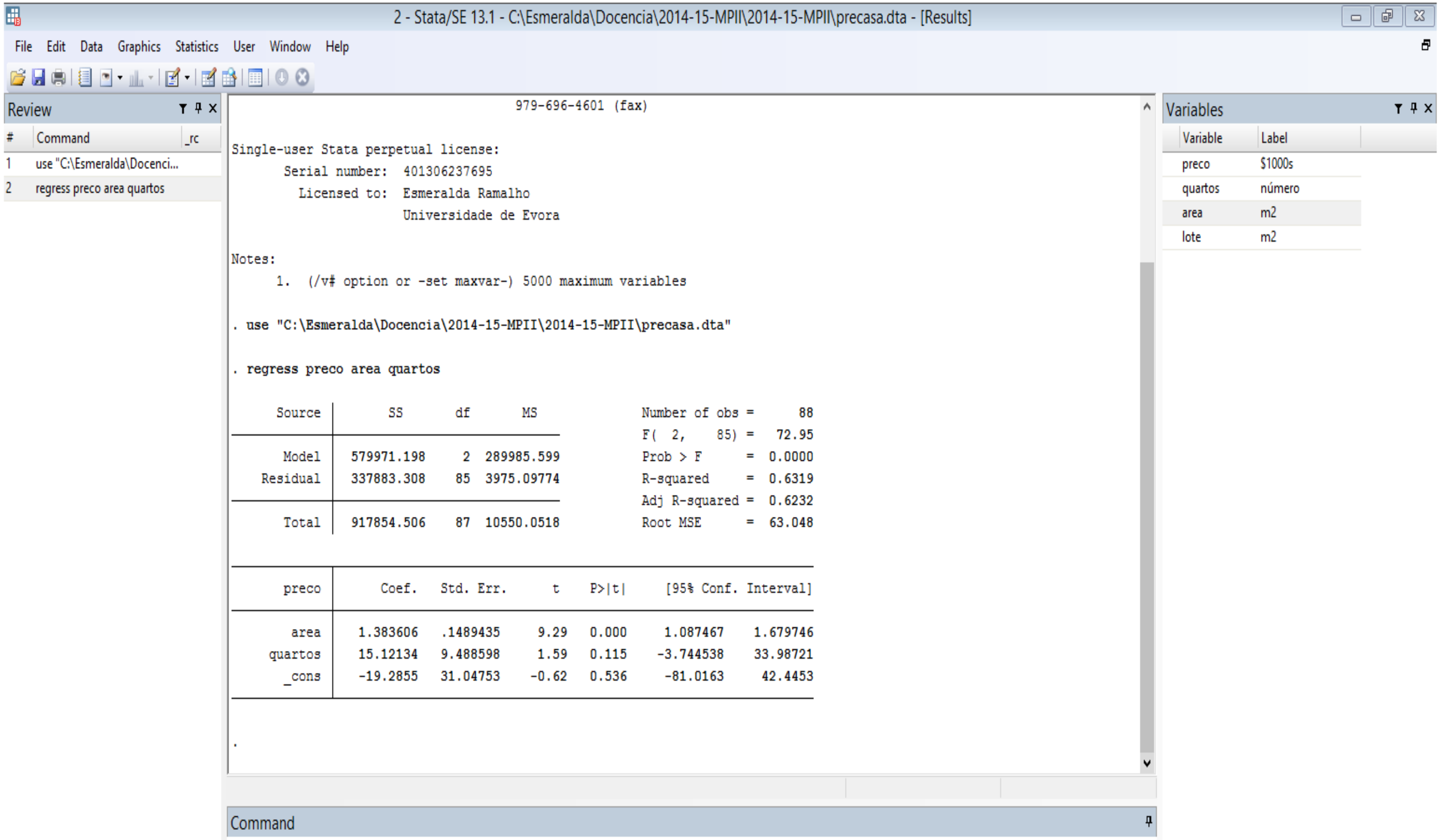
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-19.28550	31.04753	-0.621161	0.5362
AREA	1.383606	0.148943	9.289470	0.0000
QUARTOS	15.12134	9.488598	1.593632	0.1147

R-squared	0.631877	Mean dependent var	293.5460
Adjusted R-squared	0.623215	S.D. dependent var	102.7134
S.E. of regression	63.04838	Akaike info criterion	11.15918
Sum squared resid	337883.3	Schwarz criterion	11.24363
Log likelihood	-488.0038	Hannan-Quinn criter.	11.19320
F-statistic	72.95056	Durbin-Watson stat	1.857617
Prob(F-statistic)	0.000000		

Output STATA

Modelo estimado:

$$\widehat{preço} = -19.286 + 1.3836\, area + 15.121\, quartos$$




```
                              979-696-4601 (fax)

Single-user Stata perpetual license:
       Serial number:  401306237695
         Licensed to:  Esmeralda Ramalho
                       Universidade de Evora

Notes:
      1.  (/v# option or -set maxvar-) 5000 maximum variables

. use "C:\Esmeralda\Docencia\2014-15-MPII\2014-15-MPII\precasa.dta"

. regress preco area quartos

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  2,    85) =   72.95
       Model |  579971.198     2  289985.599           Prob > F      =  0.0000
    Residual |  337883.308    85  3975.09774           R-squared     =  0.6319
-------------+------------------------------           Adj R-squared =  0.6232
       Total |  917854.506    87  10550.0518           Root MSE      =  63.048

------------------------------------------------------------------------------
       preco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        area |   1.383606   .1489435     9.29   0.000     1.087467    1.679746
     quartos |   15.12134   9.488598     1.59   0.115    -3.744538    33.98721
       _cons |  -19.2855   31.04753    -0.62   0.536     -81.0163    42.4453
------------------------------------------------------------------------------

.
```

Output R

Modelo estimado:

$$\widehat{preço} = -19.286 + 1.3836\, area + 15.121\, quartos$$

```
Call:
lm(formula = preço ~ area + quartos, data = dta)

Residuals:
     Min        1Q    Median        3Q       Max
-128.056   -42.814    -7.128    32.466   229.645

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -19.2855    31.0475  -0.621    0.536
Area          1.3836     0.1489   9.289  1.4e-14 ***
quartos      15.1213     9.4886   1.594    0.115
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 63.05 on 85 degrees of freedom
Multiple R-squared:  0.6319,	Adjusted R-squared:  0.6232
F-statistic: 72.95 on 2 and 85 DF,  p-value: < 2.2e-16
```

Interpretação dos coeficientes da regressão

- A interpretação do modelo é geralmente feita em termos marginais, isto é, avaliando a reação do valor esperado condicional de y dado $\boldsymbol{x}$, $E(y|\boldsymbol{x})$, a uma variação da variável explicativa x_j, tudo o resto constante (o ***ceteris paribus***)

- Em muitas situações este impacto está diretamente relacionado com o coeficiente β_j associado com a variável explicativa x_j mas é não sempre caso. Por exemplo se o modelo for $E(y|x) = \beta_0 + \beta_1 x + \beta_2 x^2$ o impacto de uma variação de x no valor esperado condicionado será função de β_1, de β_2 e do próprio $x \rightarrow \frac{dE(y|x)}{dx} = \beta_1 + 2\,\beta_2 x$

- Comece-se então pelo caso mais simples, o modelo lin-lin, onde nenhuma variável sofreu qualquer transformação.

Caso 1: Modelo lin-lin

Modelo linear sem transformação das variáveis,

$$E(y|\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

Tem-se que

$$\Delta x_j = 1 \rightarrow \Delta E(y|\boldsymbol{x}) = \beta_j, \textit{ ceteris paribus}$$

A dedução é imediata e pode ser ilustrada para $j = 1$.

$$E(y|\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$
$$E(y|\boldsymbol{x}^*) = \beta_0 + \beta_1 (x_1 + 1) + \cdots + \beta_k x_k = E(y|\boldsymbol{x}) + \beta_1$$

com $x^* = (x_1 + 1, x_2, \dots, x_k) \rightarrow$ *ceteris paribus*

e $\Delta E(y|\boldsymbol{x}) = E(y|\boldsymbol{x}^*) - E(y|\boldsymbol{x})$

Se quer x_1 quer x_2 variarem de uma unidade:

$$\Delta x_1 = \Delta x_2 = 1 \rightarrow \Delta E(y|\boldsymbol{x}) = \beta_1 \Delta x_1 + \beta_2 \Delta x_2 = \beta_1 + \beta_2, \textit{ ceteris paribus}$$

Modelo lin-lin: exemplo

Exemplo: Considere o modelo

$$preço = \beta_0 + \beta_1 area + \beta_2 quartos + u$$

onde o **preço de uma casa**, em milhares de dólares, depende da **área** (m^2) e do **número de quartos** (dados disponíveis no FENIX).

Estimou-se

$$\widehat{preço} = -19.286 + 1.384\, area + 15.121 quartos$$

Admitindo tudo o resto constante:

- por cada m^2 adicional (variação unitária de área), é de esperar que o preço esperado da casa aumente 1384 dólares;
- por cada quarto adicional (variação unitária de quartos) é de esperar que o preço esperado da casa aumente 15121 dólares.

Modelo lin-lin: exemplo

Pode também analisar-se o impacto da variação simultánea das 2 variáveis explicativas.

o aumento estimado no preço esperado pelo acrescento de um quarto que aumenta a área da casa em 13 m² é dado por

$$1.384\Delta area + 15.121\Delta quartos = 1.384 \times 13 + 15.121 \times 1$$

ou seja 33.113 milhares de dólares

.

Tal como o modelo foi formalizado (lin-lin), as variações são todas em termos absolutos e não em termos relativos.

Transformar variáveis?

Em muitas situações práticas, pode tornar-se vantojoso formalizar o modelo de regressão linear depois de transformar uma ou mais variáveis (ou introduzindo termos de ordem superior nas variáveis x_j como por exemplo x_j^2).

Flexbiliza-se assim o modelo que se ajusta a mais situações. No exemplo anterior

$$\widehat{preço} = -19.286 + 1.384\ area + 15.121 quartos$$

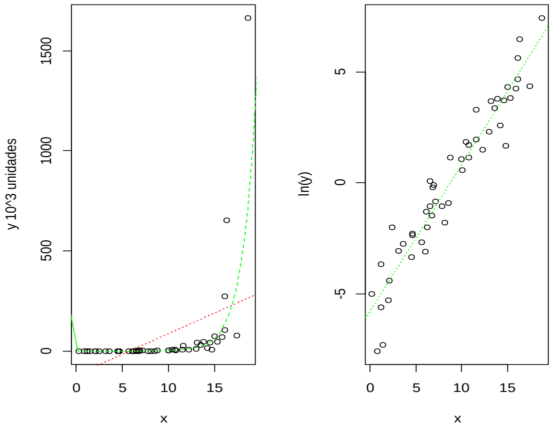
estamos a impôr que um acrescimo de 10 m2 de área tenha sempre o mesmo efeito no preço independentemente da área do imóvel o que pode não corresponder à realidade.

As transformações mais correntes passam por logaritmizar o y e/ou uma ou mais variáveis explicativas

Transformar variáveis?

Um exemplo ajuda a perceber a situação:

Depois de logaritmizar a variável y existe uma melhoria muito significativa do ajustamento



Esquerda (antes de transformar y)

Nuvem de pontos

Modelo ajustado OLS (vermelho)

Utilizando o modelo com log (verde)

Direita (y foi logaritmizado)

Nuvem de pontos

Modelo ajustado OLS (verde)

Modelos com logaritmos

Antes de avançar na interpretação dos coeficientes, duas observações:

1. A função $ln(w)$ apenas é definida para $w > 0$, por isso **não poderá ser aplicada a variáveis que assumam valores negativos ou nulos;**

2. A taxa de crescimento de uma variável w é dada por $\frac{\Delta w}{w} = \frac{w^* - w}{w}$, sendo w o valor inicial e $w^* = w + \Delta w$ o valor depois da variação Δw. Quando $\frac{\Delta w}{w}$ é pequeno,

$$\frac{\Delta w}{W} \cong ln(w + \Delta w) - ln(w)$$

Exemplo: $w = 100$ e $\Delta w = 1$, $\frac{\Delta w}{w} = \frac{101-100}{100} = 0.01$ (tx cresc. de 1%) e

$$ln(100 + 1) - ln(100) = 0.00995$$

Caso 2: Modelo log-lin

A variável y foi logaritmizada mas não a variável x

$$\mathrm{E}(ln(y)|\boldsymbol{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

ou $$ln(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

Modelo estimado: $\widehat{\ln y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$

Em termos diretos, $\Delta x_j = 1 \rightarrow \Delta \widehat{\ln y} = \hat{\beta}_j$ mas o interesse reside em $\hat{y}$ e não em $\widehat{\ln y}$.

Aplicando o resultado visto anteriormente vem

$\Delta x_j = 1 \rightarrow \%\Delta\hat{y} \cong 100\,\hat{\beta}_j\%$, *ceteris paribus*

Caso 2: Modelo log-lin

Exemplificando para $\Delta x_1 = 1$ (a que corresponde $\hat{y}^* = \hat{y} + \Delta\hat{y}$),

Modelo inicial: $\widehat{\ln y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$

Modelo pós-incremento $\widehat{\ln y^*} = \hat{\beta}_0 + \hat{\beta}_1 (x_1 + 1) + \cdots + \hat{\beta}_k x_k$

$$= \widehat{\ln y} + \hat{\beta}_1$$

$$\widehat{\ln y^*} - \widehat{\ln y} = \hat{\beta}_1 \cong \frac{\Delta\hat{y}}{\hat{y}}$$

Logo $\%\Delta\hat{y} = 100\,\hat{\beta}_1\%$ representando por $\%\Delta y$ a variação percentual de $\hat{y}$

Modelo log-lin: exemplo

Exemplo: Retomem-se os dados anteriores e considere-se agora o modelo alternativo que originou

$$\ln \widehat{preço} = 4.766 + 0.0041\, area + 0.0286\, quartos$$

Admitindo tudo o resto constante:

- Um aumento de 1 m^2 na área origina um acrescimo de aproximadamente 0.41% no preço esperado da casa
- por cada quarto adicional, o preço esperado das casas aumenta aproximadamente 2.86%

Caso 3: modelo lin-log

A variável x_1 foi logaritmizada mas não a variável y

$$\mathrm{E}(y\,|\boldsymbol{x}) = \beta_0 + \beta_1 \ln(x_1) + \cdots + \beta_k x_k$$

ou $$y = \beta_0 + \beta_1 \ln x_1 + \cdots + \beta_k x_k + u$$

Modelo estimado: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \ln x_1 + \cdots + \hat{\beta}_k x_k$

Modelo pós-incremento $\widehat{y^*} = \hat{\beta}_0 + \hat{\beta}_1 \ln(x_1 + 1) + \cdots + \hat{\beta}_k x_k$

$$\Delta\hat{y} = \widehat{y^*} - \hat{y} = \hat{\beta}_1(\ln(x_1 + 1) - \ln x_1) \cong \hat{\beta}_1 \frac{\Delta x_1}{x_1}$$

Logo $\%\Delta\, x_1 = 1 \rightarrow \Delta\hat{y} = \frac{\hat{\beta}_1}{100}$, *ceteris paribus*

Nota: como é evidente, esta dedução aplica-se para qualquer x_j

Modelo lin-log: exemplo

Exemplo: considere um novo modelo alternativo para o mesmo problema

$$\widehat{preço} = -1153.7 + 265.0\ln(area) + 19.63\ quartos$$

Admitindo tudo o resto constante:

- uma variação da área de 1% gera um aumento no preço esperado das casas de aproximadamente 2.65 milhares de doláres
- Como a função ln(.) apenas é definida para valores positivos das variáveis e nada impede que se tenha $quartos = 0$ não se logaritmizou a variável.
- Este modelo é menos utilizado do que os outros

Caso 4: modelo log-log

Quer a variável y, quer a variável x foram logaritmizadas

$$\mathrm{E}(\ln y\,|\boldsymbol{x}) = \beta_0 + \beta_1 \ln(x_1) + \cdots + \beta_k x_k$$

ou $$\ln y = \beta_0 + \beta_1 \ln x_1 + \cdots + \beta_k x_k + u$$

Modelo estimado: $\widehat{\ln y} = \hat{\beta}_0 + \hat{\beta}_1 \ln x_1 + \cdots + \hat{\beta}_k x_k$

Modelo pós-incremento $\widehat{\ln y^*} = \hat{\beta}_0 + \hat{\beta}_1 \ln(x_1 + 1) + \cdots + \hat{\beta}_k x_k$

Logo $\widehat{\ln y^*}$ - $\widehat{\ln y} = \hat{\beta}_1 (\ln(x_1 + 1) - \ln x_1)$

$$\%\Delta\hat{y} \cong \hat{\beta}_1\ \%\Delta x_1$$

Em geral, $\%\Delta\, x_j = 1 \rightarrow \%\Delta\hat{y} \cong \beta_j\%$, *ceteris paribus*

Caso 4: Modelo log-log   Exemplo

Exemplo: considere-se agora o modelo

$$\widehat{ln(preço)} = 1.289 + 0.810\, ln(area) + 0.038\, quartos$$

Admitindo tudo o resto constante:

- uma variação da área de 1% gera um aumento no preço esperado das casas de aproximadamente 0.810%
- por cada quarto adicional, o preço esperado das casas aumenta aproximadamente 3.8%

Interpretação dos coeficientes

Quadro síntese: Table 2.3 Wooldridge

y^*	x^*	Interpretação de β_j	
y	x_j	$\Delta y = \beta_j \, \Delta x_j$	
y	$\ln x_j$	$\Delta y = (\beta_j \,/100) \; \%\Delta x_j$	
$\ln y$	x_j	$\%\Delta y = (100 \, \beta_j) \; \Delta x_j$	½ elasticidade constante
$\ln y$	$\ln x_j$	$\%\Delta y = \beta_j \;\; \Delta x_j$	elasticidade constante

Aprofundando a interpretação dos $\hat{\beta}_j$

Modelo: $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$

Modelo estimado: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1\, x_1 + \cdots + \hat{\beta}_k\, x_k$ de resíduos $\hat{u}_i$

$\hat{\beta}_j$ é o elemento j de $\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\, \boldsymbol{X}^T Y$

Vamos agora analisar 2 pontos:

- Interpretação “Partialling out” de $\hat{\beta}_j$
- O que acontece a $\hat{\beta}_j$ quando se acrescenta (ou se elimina) uma das (outras) variáveis explicativas

Efeito *partialling out*

$\hat{\beta}_j$ vai medir o **efeito parcial** da variável x_j depois de considerado o efeito das restantes variáveis explicativas o que é diferente do efeito marginal que se obteria com uma regressão simples.

O teorema de Frisch-Waugh fundamenta este ponto já que prova que o $\hat{\beta}_j$ que se obtém no MRLM é o mesmo do que aquele que se obteria numa regressão linear simples de y em r_j, resíduos da regressão de x_j nas restantes variáveis explicativas.

O efeito "*partialling out*" consiste em retirar de x_j aquilo que é "comum" com as restantes varíaveis explicativas (por isso se tiram os resíduos da regressão auxiliar) antes de proceder a regressão com y.

Efeito *partialling out*

Modelo: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$

Modelo auxiliar: $x_j = \hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \cdots + \hat{\alpha}_k x_k + \hat{r}_j$ (todas menos o x_j)

RL simples: $\hat{y} = \hat{\gamma} + \hat{\beta}_j \, \hat{r}_j$

Teorema de Frisch-Waugh:

$$\hat{\beta}_j = \frac{\sum_{i=1}^{n} \hat{r}_{ij} y_i}{\sum_{i=1}^{n} \hat{r}_{ij}{}^2}, j = 1,2, \dots, k$$

onde $\hat{r}_{ij}$ são os resíduos da regressão de x_{ij} nas restantes variáveis explicativas $x_{i1}, x_{i2}, \dots x_{ik}$ (todas menos o x_{ij}). O $\hat{\beta}_j$ que se obtém é o estimador dos MQ que se obtém na regressão (com termo independente) de y_i em $\hat{r}_{ij}$.

Recorda-se que a expressão obtida para $\hat{\beta}_j$ é o estimador OLS na regressão simples.

Teorema de Frisch-Waugh - ilustração

A demonstração do teorema de Frisch-Waugh encontra-se em anexo. Vamos apenas ver uma ilustração.

Retome-se o exemplo e estime-se a regressão de $area_i$ em $quartos_i$

$$\widehat{preço} = -19.286 + 1.384\, area + 15.121\, quartos$$

.

Pode obter-se $\hat{\beta}_1$ fazendo:

Regressão de $area$ em $quartos$ designando por $\hat{r}_1$ os resíduos desta regressão

$$\widehat{area} = -19.286 + 33.899\, quartos$$

Regressão de $preço$ em $\hat{r}_1$

$$\widehat{preço} = 293.546 + 1.384\,\hat{r}_1$$

Teorema de Frisch-Waugh - ilustração

Regressão original

$$\widehat{preço} = -19.286 + \mathbf{1.3836}\ area + 15.121\ quartos$$

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.794906945					
R Square	0.63187705					
Adjusted R Square	0.623215334					
Standard Error	63.04837628					
Observations	88					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	2	579971.1994	289985.5997	72.95055817	3.58672E-19	
Residual	85	337883.3088	3975.097751			
Total	87	917854.5083				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-19.28550028	31.0475285	-0.621160563	0.536156232	-81.0163048	42.44530424
area(m2)	1.38360615	0.148943494	9.289470184	1.40049E-14	1.08746658	1.67974572
quartos	15.12133684	9.488597692	1.593632413	0.114730383	-3.744537442	33.98721112

Teorema de Frisch-Waugh - ilustração

1ª regressão auxilar: $area$ como função de $quartos$

$$\widehat{area} = 66.1435 + 33.8993quartos$$

Os $\hat{r}_{i1}$ serão os resíduos desta regressão

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.532120204					
R Square	0.283151911					
Adjusted R Square	0.274816468					
Standard Error	45.64604375					
Observations	88					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	1	70777.80685	70777.80685	33.96963003	9.52941E-08	
Residual	86	179186.2727	2083.56131			
Total	87	249964.0795				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	66.14354244	21.31632029	3.102953114	0.002592652	23.76810017	108.5189847
quartos	33.89926199	5.816273698	5.828347109	9.52941E-08	22.33689256	45.46163143

Teorema de Frisch-Waugh - ilustração

2ª regressão auxiliar $preço$ como função de $\hat{r}_1$ (resid)

$$\widehat{preço} = 293.546 + \mathbf{1.3836}\,\hat{r}_1$$

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.611333125					
R Square	0.37372819					
Adjusted R Square	0.36644596					
Standard Error	81.75590471					
Observations	88					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	1	343028.1042	343028.1042	51.32056696	2.5121E-10	
Residual	86	574826.4041	6684.027955			
Total	87	917854.5083				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	293.5460341	8.715208725	33.68204289	2.57951E-51	276.2207726	310.8712956
resid	1.38360615	0.193137569	7.163837446	2.5121E-10	0.99966137	1.76755093

Efeito da adição de uma variável

Quando se acrescenta uma variável explicativa no modelo, os coeficientes das restantes variáveis alteram-se a não ser em dois casos muito especiais:

- O novo regressor é ortogonal em relação a qualquer dos outros
- O impacto do novo regressor é nulo

Claro que estas conclusões também são válidas (depois de adaptadas) quando se trata da eliminação de uma variável explicativa.

Efeito da adição de uma variável

Este ponto torna-se claro ao mostrar que (ver Wooldridge ou demonstração em anexo)

$$\tilde{\beta}_j = \hat{\beta}_j + \hat{\beta}_r\, \tilde{\delta}_j \quad j, r = 1,2, \dots, k \text{ e } r \neq j$$

$\hat{\beta}_j$ coeficiente de x_j na regressão de y em $\boldsymbol{x} = (x_1, x_2, \dots, x_k)$.

Inclui todas as variáveis explicativas

$\tilde{\beta}_j$ coeficiente de x_j na regressão de y em $\boldsymbol{x}$ com exceção de x_r.

$\tilde{\delta}_j$ coeficiente de x_j na regressão de x_r nas restantes variáveis explicativas, isto é, $\boldsymbol{x}$ com exceção de x_r.

Logo, $\tilde{\beta}_j = \hat{\beta}_j$ vai implicar $\hat{\beta}_r = 0$ (a variável x_r não tem impacto em y) ou $\tilde{\delta}_j = 0$ (ortogonalidade entre x_r e x_j)

Efeito da adição de uma variável - ilust

Exemplo (dados no Wooldridge – wage1)

Modelo: salário/hora em dolares; educação, tenure (nº anos na empresa) e exper (experiência profissional) em anos; female=1 se mulher e 0 se homem.

$$\widehat{wage.h} = -0.84503 + \mathbf{0.53799} educ + 0.16441\, tenure - 1.78839\, female$$

Qual o efeito no coeficiente de $educ$ **ao introduzir a variável** $exper$?

Em termos práticos estima-se uma nova regressão

$$\widehat{wage.h} = -1.56794 + \mathbf{0.57150} educ + 0.14101\, tenure - 1.81085\, female + 0.02539\, exper$$

E calcula-se $\hat{\beta}_1 - \tilde{\beta}_1 = 0.57150 - 0.53799 = 0.03351$

Para verificar o resultado teórico, estima-se um terceira regressão

$$\widehat{exper} = 28.4654 - \mathbf{1.31952} educ + 0.92169\, tenure + 0.88433\, female$$

e calcula-se $-\hat{\beta}_4 \tilde{\delta}_1 = -0.02539 \times (-1.31952) = 0.03350$ (arredondamento)

A variação (relativamente fraca) fica-se a dever ao efeito parcial de $exper$ no salario (fraco) mas também no efeito parcial de $educ$ em exper (mais forte).

Efeito da adição de uma variável - ilust

$$\widehat{wage.h} = -0.84503 + \mathbf{0.53799}\ educ + 0.16441\ tenure - 1.78839\ female$$

SUMMARY OUTPUT			wage.h			
Regression Statistics						
Multiple R	0.598042764					
R Square	0.357655148					
Adjusted R Square	0.353963511					
Standard Error	2.968369417					
Observations	526					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	3	2560.959037	853.6530124	96.88253196	7.36418E-50	
Residual	522	4599.455273	8.811216998			
Total	525	7160.41431				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-0.845034444	0.647737253	-1.30459448	0.192605839	-2.117526547	0.427457658
educ.years	0.537994321	0.047087329	11.42545852	3.87633E-27	0.445490372	0.63049827
tenure.years	0.164412254	0.018345337	8.96207322	5.6557E-18	0.128372492	0.200452016
female	-1.788393867	0.265593604	-6.733572799	4.377E-11	-2.310157533	-1.266630201

Efeito da adição de uma variável - ilust

$$\widehat{wage.h} = -1.56794 + \mathbf{0.57150}\ educ + 0.14101\ tenure - 1.81085\ female$$
$$+0.02539\ exper$$

SUMMARY OUTPUT			wage.h			
Regression Statistics						
Multiple R	0.602943893					
R Square	0.363541338					
Adjusted R Square	0.358654899					
Standard Error	2.957571925					
Observations	526					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	4	2603.106599	650.7766497	74.39801215	7.2956E-50	
Residual	521	4557.307712	8.747231692			
Total	525	7160.41431				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-1.567938772	0.724551117	-2.164014016	0.030916969	-2.991339514	-0.14453803
educ.years	0.571504775	0.049337316	11.58362113	9.09319E-28	0.474580251	0.668429299
tenure.years	0.141005064	0.021161685	6.663224735	6.82937E-11	0.099432347	0.182577781
female	-1.810852171	0.264825211	-6.837914575	2.25707E-11	-2.331108636	-1.290595707
exper.years	0.025395866	0.011569434	2.195082824	0.028598072	0.002667392	0.048124339

Efeito da adição de uma variável - ilust

$$\widehat{exper} = 28.4654 - \mathbf{1.31952}educ + 0.92169\ tenure + 0.88433\ female$$

SUMMARY OUTPUT		exper.years				
Regression Statistics						
Multiple R	0.569425246					
R Square	0.324245111					
Adjusted R Square	0.320361462					
Standard Error	11.18891453					
Observations	526					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	3	31356.72199	10452.24066	83.48981289	3.92223E-44	
Residual	522	65350.12402	125.1918085			
Total	525	96706.84601				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	28.46543377	2.44156833	11.65866768	4.46944E-28	23.66892654	33.261941
educ.years	-1.319524009	0.17749007	-7.434353978	4.3351E-13	-1.668206614	-0.970841404
tenure.years	0.921692939	0.06915056	13.32878484	4.31713E-35	0.785845354	1.057540525
female	0.88432915	1.001123417	0.883336794	0.377460959	-1.082396766	2.851055067

Hipótese MRL 1

Para analisar a qualidade do estimador OLS e para poder definir procedimentos de inferência estatística torna-se necessário assumir um conjunto de hipóteses sobre o MRLM.

MLR 1 – Modelo linear nos parâmetros

O "verdadeiro" modelo na população é dado por

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

onde β é um vetor de parâmetros e u um termo de erro aleatório e não observável.

Observações:

1. Caso a relação entre y e $\boldsymbol{x}$ na população não seja esta, estariamos a utilizar um modelo "errado"
2. Tenha-se presente a **flexibilidade** "adicional" dada pela posibilidade de "**transformar"** as variáveis.

Hipótese MRL 2

MLR 2 – Amostra aleatória

As n observações $(y_i, x_{i1}, x_{i2}, \dots, x_{ik})$, $i = 1,2, \dots, n$, constituem uma amostra aleatória retirada da população.

Observações:

1. O modelo definido na hipótese MLR 1 aplica-se a qualquer das observações, $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i$. O índice i refere-se às observações ($i = 1,2, \dots, n$) e o índice j, quando necessário, às variáveis explicativas ($j = 1,2, \dots, k$).
2. As observações serão consideradas independentes (modelo seccional).

Hipótese MRL 3

MLR 3 – Ausência de colinearidade perfeita

Nenhuma das variáveis explicativas pode ser combinação linear das restantes nem na amostra nem na população.

Observações:

1. Esta hipótese destina-se a garantir que a matriz $\boldsymbol{X}^T\boldsymbol{X}$ tem inversa, isto é, que $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ existe. Recorda-se que $\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T Y$ e portanto que a existência de $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ é necessária para que o estimador $\hat{\beta}$ exista.
2. Quando se verificar colinearidade perfeita na amostra esta será facilmente identificável no output.
3. **Importante** – Estamos a impor a **não exitência de colinearidade perfeita** mas pode existir correlação entre as variáveis explicativas. Quando esta atingir valores "significativos" teremos problemas de **multicolinearidade** a serem abordados mais adiante.

Hipótese MRL 4

MLR 4 – Exogeneidade

$$E(u|\boldsymbol{x}) = 0$$

Observações:

1. Esta hipótese é **fundamental.** O ponto essencial é que não exista correlação entre as variáveis explicativas x_j e o termo de erro u. Caso u esteja correlacionado com uma variável explicativa esta diz-se endógena e a hipótese MLR 4 … falha!
2. Ela pode falhar por má especificação do modelo: omissão de variáveis explicativas correlacionadas com as restantes ou não transformação (quando adequada) de variáveis. Problemas a serem anlisados mais adiante (eventualmente para além do programa da UC)
3. Enquanto MRL 3 diz respeito apenas às variáveis explicativas incluídas (nada tem a ver com o termo de erro u), MLR 4 aborda as variáveis não incluídas que se encontram “representadas” por u.

$\hat{\beta}$ estimador centrado de β

Verificadas as hipóteses MRL 1 a MLR 4, prova-se que $E(\hat{\beta}_j) = \beta_j, j = 1,2, \dots, k$.

Demonstração:

$Y = \boldsymbol{X}\beta + U$ (MRL 1), $\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T Y$ (MRL 2 e 3), logo

$$\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{X}\beta + U) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\beta + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T U$$
$$= \beta + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T U$$

Assim

$$E(\hat{\beta}|\boldsymbol{X}) = E(\beta|\boldsymbol{X}) + E\left((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T U|\boldsymbol{X}\right) = \beta + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T E(U|\boldsymbol{X})$$
$$= \beta \qquad \text{(MRL 4)}$$

Logo $E(\hat{\beta}) = E\left(E(\hat{\beta}|\boldsymbol{X})\right) = E(\beta) = \beta$

sendo, por isso $\hat{\beta}$ um estimador centrado

Hipótese MRL 5

MLR 5 – Homocedasticidade

$$var(u|\boldsymbol{x}) = \sigma^2$$

Observações:

1. A variância da variável residual u é constante, isto é, não depende dos valores assumidos pelas variáveis explicativas.
2. Esta hipótese nem sempre é muito realista e o seu levantamento será discutido mais adiante.
3. Tendo em conta que as observações da amostra são independentes (bastaria não correlacionadas), $var(U|\boldsymbol{X}) = \sigma^2\, I$, isto é,

   $var(u_i|\boldsymbol{X}_{i\cdot}) = \sigma^2$ e $cov\left(u_i, u_j\middle|\boldsymbol{X}\right) = 0,\ i, j = 1,2, \dots, n,\ i \neq j$

Hipótese MRL 5 - ilustração

FIGURE 2.8

The simple regression model under homoskedasticity.

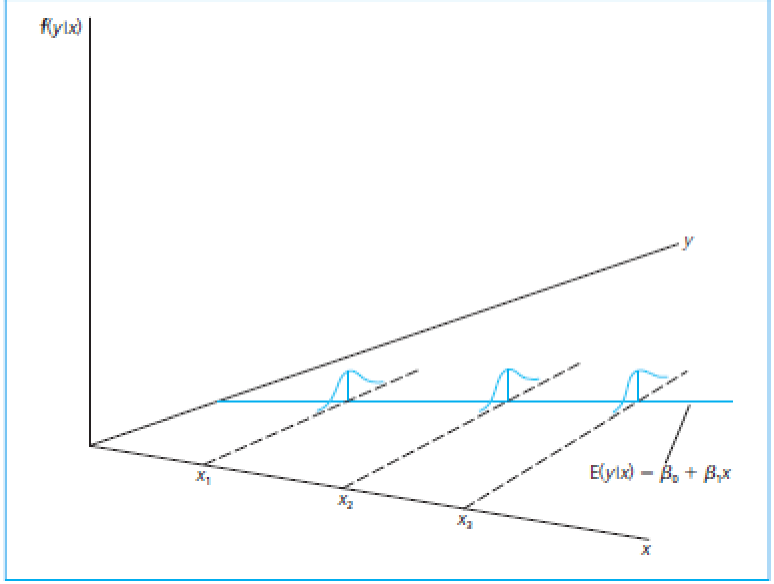
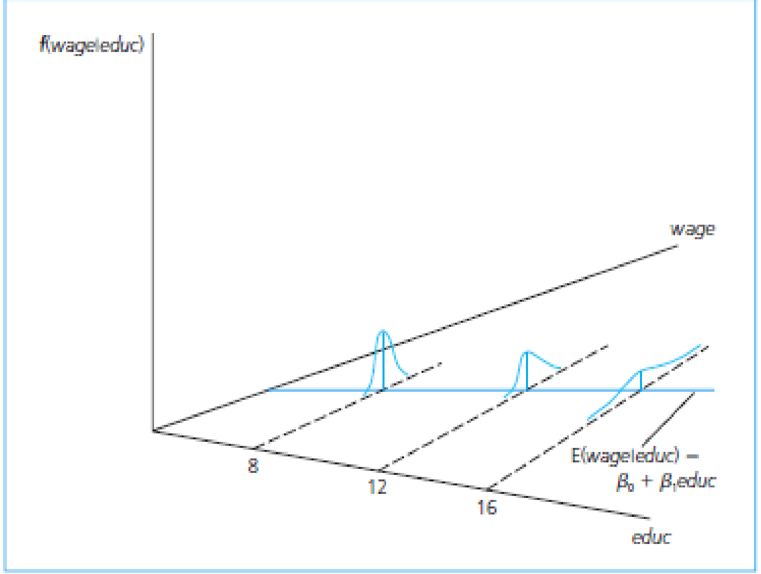


FIGURE 2.9

Var(wage|educ) increasing with educ.



Matriz das var/cov do estimador OLS

Recordando-se que $\hat{\beta} = \beta + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T U$ (slide 5 desta série) e que $var(U|\boldsymbol{X}) = \sigma^2 I$, pode-se obter a matriz de variâncias e covariâncias de $\hat{\beta}$.

$$
\begin{aligned}
var(\hat{\beta}|\boldsymbol{X}) &= var[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T U|\mathbf{X}] \qquad (\beta \text{ é não aleatório})\\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T var(U|\boldsymbol{X})\, \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\sigma^2\, \boldsymbol{I})\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\\
&= \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\\
&= \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}
\end{aligned}
$$

Como σ^2 é um parâmetro desconhecido, $var(\hat{\beta}|\boldsymbol{X})$ será, ela também desconhecida. Para estimar $var(\hat{\beta}|\boldsymbol{X})$, matriz simétrica de dimensão $\left((k+1)\times(k+1)\right)$, bastará estimar σ^2 uma vez que a matriz $\boldsymbol{X}$ é observável.

Matriz das var/cov do estimador OLS

A variância condicionada de um dado $\hat{\beta}_j$, $j = 1,2,\dots,k$, que se designa por $\sigma^2_{\hat{\beta}_j}$, é dada pelo elemento (j,j) na diagonal principal da matriz $\sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$, enquanto a covariância condicionada entre $\hat{\beta}_i$ e $\hat{\beta}_r$ será dada pelo elemento (i,r) da matriz ou pelo elemento (r,i) já que $cov(\hat{\beta}_i, \hat{\beta}_r) = cov(\hat{\beta}_r, \hat{\beta}_i)$.

$$var(\hat{\beta}|\boldsymbol{X}) = \begin{bmatrix} var(\hat{\beta}_0|\boldsymbol{X}) & cov(\hat{\beta}_0, \hat{\beta}_1|\boldsymbol{X}) & \dots \\ cov(\hat{\beta}_1, \hat{\beta}_0|\boldsymbol{X}) & var(\hat{\beta}_1|\boldsymbol{X}) & \dots \\ \dots & & \dots \\ cov(\hat{\beta}_k, \hat{\beta}_0|\boldsymbol{X}) & cov(\hat{\beta}_k, \hat{\beta}_1|\boldsymbol{X}) & \end{bmatrix}$$

Cada elemento desta matriz é dado pelo produto de σ^2 (desconhecido) pelo correspondente elemento da matriz $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$

Estimação de σ^2

Para estimar σ^2 utiliza-se

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i{}^2}{n-k-1} = \frac{\hat{U}^T \hat{U}}{n-k-1} = MSE$$

que é estimador **centrado** de σ^2.

Observação: $\hat{\sigma}$ é designado por **erro padrão da regressão.**

Porquê este estimador?

Sendo $\sigma^2 = var(u_i|\boldsymbol{x})$ a solução "natural" seria considerar a variância observada de u, $S_u^2 = \frac{\sum_{i=1}^{n}(u_i - \bar{u})^2}{n}$ ou a sua variância corrigida. Mas, não sendo u observável, tais soluções não são possíveis.

A ideia é então utilizar $\hat{u}_i$ em vez de u_i, introduzindo-se uma compensação no denominador destinada a garantir que o estimador é centrado.

A demonstração de que $E(\hat{\sigma}^2) = \sigma^2$ está fora do âmbito do curso

Matriz estimada das var/cov de $\hat{\beta}$

A matriz estimada das variâncias/covariâncias condicionais de $\hat{\beta}$ será então

$$\widehat{var}(\hat{\beta}|\boldsymbol{X}) = \hat{\sigma}^2(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{bmatrix} \widehat{var}(\hat{\beta}_0|\boldsymbol{X}) & \widehat{cov}(\hat{\beta}_0,\hat{\beta}_1|\boldsymbol{X}) & \cdots \\ \widehat{cov}(\hat{\beta}_1,\hat{\beta}_0|\boldsymbol{X}) & \widehat{var}(\hat{\beta}_1|\boldsymbol{X}) & \cdots \\ \cdots & & \cdots \\ \widehat{cov}(\hat{\beta}_k,\hat{\beta}_0|\boldsymbol{X}) & \widehat{cov}(\hat{\beta}_k,\hat{\beta}_1|\boldsymbol{X}) & \end{bmatrix}$$

Tal como $\hat{\sigma}$ é designado por erro padrão da regressão, define-se o erro padrão (standard error) do estimador $\hat{\beta}_j$ como $\hat{\sigma}_{\hat{\beta}_j} = \sqrt{\widehat{var}(\hat{\beta}_j|\boldsymbol{X})}$

No caso particular do MRL simples $y = \beta_0 + \beta_1 x + u$ obtém-se

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Modelo estimado:

$$\widehat{preço} = -19.286 + \mathbf{1.3836}\, area + 15.121\, quartos$$

Identificar:

$$\hat{\beta} = \begin{bmatrix} -19.286 \\ 1.3836 \\ 15.121 \end{bmatrix}$$

$$\hat{\sigma}_{\hat{\beta}_0} = 31.048$$
$$\hat{\sigma}_{\hat{\beta}_1} = 0.1489$$
$$\hat{\sigma}_{\hat{\beta}_2} = 9.4886$$

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.794906945					
R Square	0.63187705					
Adjusted R Square	0.623215334					
Standard Error	63.04837628					
Observations	88					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	2	579971.1994	289985.5997	72.95055817	3.58672E-19	
Residual	85	337883.3088	3975.097751			
Total	87	917854.5083				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-19.28550028	31.0475285	-0.621160563	0.536156232	-81.0163048	42.44530424
area(m2)	1.38360615	0.148943494	9.289470184	1.40049E-14	1.08746658	1.67974572
quartos	15.12133684	9.488597692	1.593632413	0.114730383	-3.744537442	33.98721112

Output R

Output Principal

```
Call:
lm(formula = preço ~ area + quartos, data = dta)

Residuals:
     Min       1Q   Median      3Q     Max
-128.056  -42.814   -7.128  32.466  229.645

Coefficients:
            Estimate  Std. Error  t value  Pr(>|t|)
(Intercept) -19.2855    31.0475   -0.621   0.536
area          1.3836     0.1489    9.289  1.4e-14 ***
quartos      15.1213     9.4886    1.594   0.115
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 63.05 on 85 degrees of freedom
Multiple R-squared:  0.6319,    Adjusted R-squared:  0.6232
F-statistic: 72.95 on 2 and 85 DF,  p-value: < 2.2e-16
```

Output adicionais R

Matriz das var/cov estimadas

```
> vcov(out1)
             (Intercept)          area       quartos
(Intercept)   963.949026   -1.46733923  -180.5499241
area           -1.467339    0.02218416    -0.7520268
quartos      -180.549924   -0.75202680    90.0334862
```

Matrizes XTX e XTX inversa

$X^T X$			$(X^T X)^{-1}$		
88	16465	314	0.2424969	-0.0003691	-0.0454202
16465	3330603	60838	-0.0003691	5.5808E-06	-0.0001892
314	60838	1182	-0.0454202	-0.0001892	0.0226494

Output Eviews

Modelo estimado:

$$\widehat{preço} = -19.286 + 1.3836\, area + 15.121\, quartos$$

Identificar: $R^2 = 0.6319$ $\bar{R}^2 = 0.6232$ $n = 88$ $SSR = \sum \hat{u}_i^2 = 337883.3$

$$\hat{\beta} = \begin{bmatrix} -19.286 \\ 1.3836 \\ 15.121 \end{bmatrix}$$

Dependent Variable: PRECO				
Method: Least Squares				
Date: 03/24/20 Time: 17:00				
Sample: 1 88				
Included observations: 88				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-19.28550	31.04753	-0.621161	0.5362
AREA	1.383606	0.148943	9.289470	0.0000
QUARTOS	15.12134	9.488598	1.593632	0.1147
R-squared	0.631877	Mean dependent var		293.5460
Adjusted R-squared	0.623215	S.D. dependent var		102.7134
S.E. of regression	63.04838	Akaike info criterion		11.15918
Sum squared resid	337883.3	Schwarz criterion		11.24363
Log likelihood	-488.0038	Hannan-Quinn criter.		11.19320
F-statistic	72.95056	Durbin-Watson stat		1.857617
Prob(F-statistic)	0.000000			

Output Eviews

Coefficient Covariance Matrix

	C	AREA_M2_	QUARTOS
C	963.9490	-1.467339	-180.5499
AREA	-1.467339	0.022184	-0.752027
QUART	-180.5499	-0.752027	90.03349

Teorema de Gauss Markov – Estimador BLUE

Verificadas as hipóteses MRL 1 a MRL 5, prova-se (**teorema de Gauss - Markov**) que, de entre os estimadores lineares e centrados, o estimador OLS é o de menor variância, sendo assim o mais eficiente → demonstração no Wooldridge.

Assim, o estimador OLS é BLUE (Best Linear Unbiased Estimator)

Best → Menor variância

Linear → Cada $\hat{\beta}_j$ é uma combinação linear dos y_i. O estimador é assim linear em y.

$\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T Y = \boldsymbol{A}\, Y$ em que $\boldsymbol{A} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ é uma matriz $\big((k+1)\times n\big)$. $\hat{\beta}_j = \sum_{i=1}^{n} a_{i,j}\, y_i$ sendo $a_{i,j}$ o elemento (i,j) da matriz $\boldsymbol{A}$.

"**Unbiased**" → $E\big(\hat{\beta}\big) = \beta$ como se mostrou anteriormente.

Demonstração: ver anexo ao capítulo 3 no Wooldridge

Multicolinearidade

Utilizando o teorema de Frisch–Waugh, é possível aprofundar a interpretação de $var(\hat{\beta}_j|\boldsymbol{X})$ já que

$$var(\hat{\beta}_j|\boldsymbol{X}) = \frac{\sigma^2}{SST_j\,(1 - R_j^2)}$$

Em que $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$ e R_j^2 o coeficiente de determinação da regressão de x_j nas restantes variáveis explicativas. Demonstração no apêndice do Wooldridge

Ilustração com base no exemplo anterior (utilizando $\hat{\sigma}^2$):

- Do modelo principal tira-se $\hat{\sigma}_{\hat{\beta}_1} = 0.148943494$ e $\hat{\sigma}^2 = 3975.097751$
- Da regressão auxiliar tira-se $SST_1 = 249964.0795$ e $R_1^2 = 0.283151911$
- Como se pode verificar, $0.148943494^2 = \frac{3975.097751}{249964.0795 \times 0.283151911}$

Multicolinearidade

$$var\left(\hat{\beta}_j \middle| \boldsymbol{X}\right) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)}$$

Comentários:

- σ^2 é a variância da variável residual. Quanto menor esta variância for, melhor o modelo e portanto menos a variabilidade dos estimadores. Note-se que este elemento, ao contrário dos outros 2, não depende de j, isto é, é comum a todos os $\hat{\beta}_j$
- SST_j é a variação total da variável x_j na amostra. Quanto maior for esta variação, menor virá a variância condicionada de $\hat{\beta}_j$. Note-se que como $SST_j = \sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2$ uma maior dimensão da amostra origina um crescimento de SST_j.
- R_j^2 . Quando mais dependente estiver x_j das restantes variáveis explicativas, maior será o valor de R_j^2 e consequentemente maior será a variância condicionada de $\hat{\beta}_j$. Esta componente merece uma análise mais detalhada.

Multicolinearidade

$$var(\hat{\beta}_j|\boldsymbol{X}) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

A melhor situação (que quase nunca acontece) é quando $R_j^2 = 0$, isto é quando x_j é ortogonal a qualquer das outras variáveis explicativas.

No extremo oposto se $R_j^2 = 1$, viola-se MLR 3 e $var(\hat{\beta}_j|\boldsymbol{X})$ vem infinita.

A multicolinearidade surge quando R_j^2 é grande, sem atingir o valor limite de 1. Neste caso, existe o perigo de $var(\hat{\beta}_j|\boldsymbol{X})$ se encontrar inflacionada, embora um elevado valor de SST_j possa mitigar (parcialmente pelo menos) este problema.

De qualquer forma, o estimador continua a ser BLUE embora o estimador tenha fraca qualidade.

Multicolinearidade

Possíveis soluções:

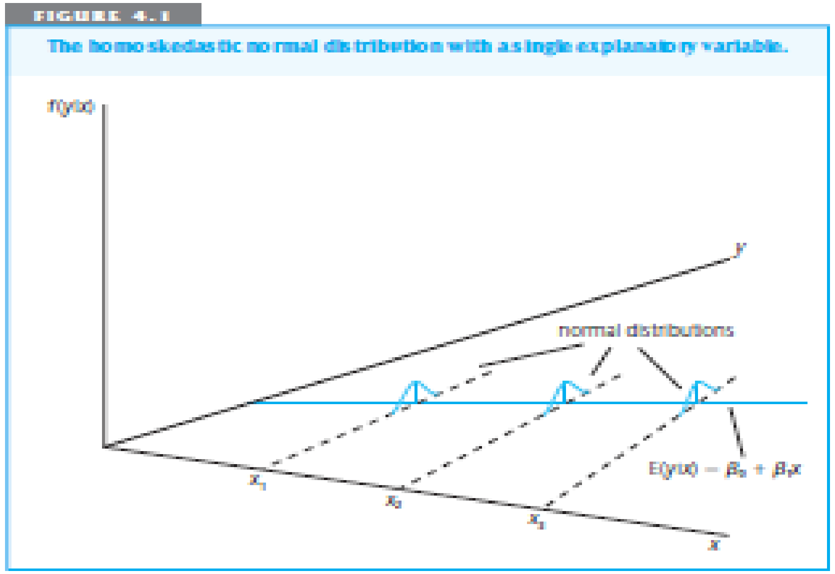
- Recolher mais informação (nem sempre funciona se esta não quebrar a excessiva dependência)

- Introduzindo alguma restrição linear entre os β_j que permita reformular o modelo: Que restrição? Não esquecer que se trata de algo que tem de se verificar para a população

- Eliminar alguma variável explicativa: Muito cuidado já que, como se irá ver, é muito mais grave eliminar erradamente uma variável explicativa do que incluir (erradamente) uma variável adicional.

Hipótese MLR 6 – Distribuição normal

O termo de erro u é independente das variáveis explicativas $x_1, x_2, \ldots, x_k$ e tem **distribuição normal** de média 0 e variância σ^2



$$y|\boldsymbol{x} \sim n(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k;\ \ \sigma^2)$$

Hipótese MLR 6

Observações:

1. Esta hipótese é bastante mais forte do que qualquer das anteriores já que, para além da distribuição normal, ela implica MLR 4 e MLR 5.

   A independência entre u e $\boldsymbol{x}$ implica $E(u|\boldsymbol{x}) = E(u)$ e $\mathrm{var}(u|\boldsymbol{x}) = var(u)$ logo (2ª parte da hipótese) $E(u) = 0$ e $\mathrm{var}(u) = \sigma^2$.

2. O conjunto de hipóteses MLR 1 a MLR 6 é conhecido como "classical linear assumptions" (**CLM**) no quadro dos modelos seccionais.

3. No quadro das hipóteses CLM o estimador $\hat{\beta}$ tem propriedades reforçadas: De estimador BLUE passa a ser o mais eficiente de entre os estimadores centrados (a comparação pode ser feita abrangendo estimadores não lineares)

Distribuição por amostragem do estimador OLS

Para as n observações da amostra,

$$y_i|\boldsymbol{x}_i \sim n(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik};\ \ \sigma^2)$$

Logo, em termos matriciais, $\mathrm{Y}|\mathbf{X} \sim N(\boldsymbol{X}\beta, \sigma^2 I)$.

Como $\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T Y$ é uma combinação linear de normais, $\hat{\beta}$ terá distribuição normal e portanto $\hat{\beta} \sim N(\beta, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$.

Para cada $\hat{\beta}_j$ vem $\hat{\beta}_j \sim N\left(\beta_j, \sigma^2_{\hat{\beta}_j}\right)$

Estandardizando $\hat{\beta}_j$, tem-se

$$\frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \sim N(0,1)$$

Inferência sobre um parâmetro β_j

A expressão $\frac{\hat{\beta}_j-\beta_j}{\sigma_{\hat{\beta}_j}} \sim N(0,1)$ não pode ser diretamente utilizada para fazer inferência sobre β_j porque envolve $\sigma_{\hat{\beta}_j}$ que é desconhecido.

A ideia é utilizar o estimador de $\sigma^2_{\hat{\beta}_j}$ em vez do parâmetro desconhecido. Mostra-se que

$$t_j = \frac{\hat{\beta}_j-\beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \dot{\sim} t(n-k-1)$$

resultado com base no qual se fará inferência (intervalos de confiança, teste de hipóteses) sobre β_j

Inferência sobre um parâmetro β_j

A intuição por trás do resultado $t_j = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t(n-k-1)$ pode ser dada comparando com o procedimento adotado em Estatística 1

Estatística 1	**MRLM**
$Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0;1)$	$\dfrac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \sim N(0{,}1)$
T= $\frac{\bar{X}-\mu}{S'/\sqrt{n}} \sim t(n-1)$	$t_j = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t(n-k-1)$
Já que $\frac{(n-1)S'^2}{\sigma^2} \sim \chi^2_{(n-1)}$	E também se poderia provar que $\frac{(n-k-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-k-1)}$

Intervalo de confiança para β_j

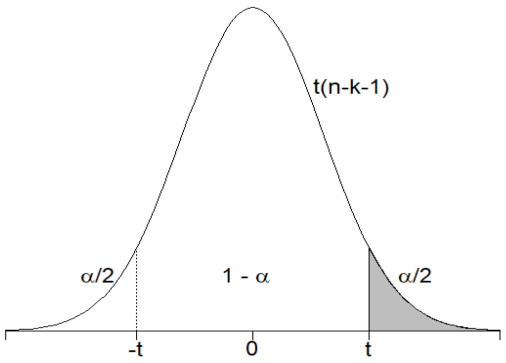
Tendo em conta o rácio t

$$t_j = \frac{\widehat{\beta}_j - \beta_j}{\widehat{\sigma}_{\widehat{\beta}_j}} \sim t(n-k-1)$$

O intervalo de confiança (com $(1-\alpha)100\%$ de confiança) para β_j vem

$$\left(\hat{\beta}_j - t_{\alpha/2}\hat{\sigma}_{\widehat{\beta}_j};\ \widehat{\beta}_j + t_{\alpha/2}\hat{\sigma}_{\widehat{\beta}_j}\right)$$

Sendo $t_{\alpha/2}$ o quantil adequado de uma distribuição *t-Student* com $(n-k-1)$ graus de liberdade.



Intervalo de confiança para β_j - exemplo

Exemplo: Considere-se o exemplo que nos tem acompanhado e que originou (erro padrão por baixo de cada coeficiente)

$$\widehat{preço} = -19.286 + 1.384\, area + 15.121\, quartos$$
(31.046) (0.1489) (9.4886)

Queremos construir um IC a 95% para o impacto da área, isto é, para β_1.

Como $n = 88$ e $k = 2$ a *t-Student* terá 85 graus de liberdade.

$t_{\alpha/2} = t_{0.025} = 1.9883$ (computador). Pela tabela fazendo a média dos valores para 80 e 90 graus de liberdade dá 1.9885.

$t_{\alpha/2}\,\hat{\sigma}_{\hat{\beta}_j} = 1.9883 \times 0.1489 = 0.296$

O IC vem então $(1.384 - 0.296; 1.384 + 0.296)$ ou seja (1.088; 1.680)

Intervalo de confiança para β_j - exemplo

Repare-se que no Excel este IC vem dado (não esquecer o efeito arredondamento nas contas que fizemos)

Output EXCEL

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.794906945					
R Square	0.63187705					
Adjusted R Square	0.623215334					
Standard Error	63.04837628					
Observations	88					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	2	579971.1994	289985.5997	72.95055817	3.58672E-19	
Residual	85	337883.3088	3975.097751			
Total	87	917854.5083				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-19.28550028	31.0475285	-0.621160563	0.536156232	-81.0163048	42.44530424
area(m2)	1.38360615	0.148943494	9.289470184	1.40049E-14	1.08746658	1.67974572
quartos	15.12133684	9.488597692	1.593632413	0.114730383	-3.744537442	33.98721112

Teste de hipóteses para β_j

Para efetuar um teste de hipóteses sobre o parâmetro β_j constrói-se a estatística de teste a partir do rácio t.

$$t_j = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t(n-k-1)$$

O procedimento é em tudo semelhante ao que seguiu no capítulo do testes de hipóteses para a média de uma população normal de variância desconhecida.

Neyman-Pearson	Valor-p
Estabelecer as hipóteses em teste, H_0 e H_1	
Definir a estatística de teste (substituir β_j pelo valor (fronteira) em H_0)	
Definir α	Calcular $t_{j,obs}$
Construir W (região de rejeição)	Calcular o valor-p
Calcular $t_{j,obs}$	Decidir
Decidir	

Teste de hipóteses para β_j - Exemplo

Exemplo: Suponha-se que, no exemplo anterior, se quer testar se um acréscimo de 10 m2 na área tem um impacto esperado superior a 10 mil dólares.

$$\widehat{preço} = -19.286 + 1.384\, area + 15.121\, quartos$$
$$\quad (31.046) \quad (0.1489) \qquad (9.4886)$$

$H_0: \beta_1 \leq 1.0$ vs $H_1: \beta_1 > 1.0$ (10 mil dólares por 10 m2 ↔ 1000 dolares/m2)

Estatística teste: $T = \frac{\hat{\beta}_1 - 1.0}{\hat{\sigma}_{\hat{\beta}_1}} \sim t(85)$

N-P: $\alpha = 0.05$ (por hip) logo $t_\alpha = t_{0.05} = 1.663$ e $W = \{t: t > 1.663\}$

$t_{1,obs} = \frac{1.384 - 1.0}{0.1489} = 2.579$ logo rejeita-se H_0, isto é, o impacto é sup. 10 mil dólares.

Valor-*p:* $P(T \geq 2.579) = 0.0058$ e tira-se a mesma conclusão

Significância de x_j

De entre os possíveis valores de β_j, o valor 0 tem uma significado particularmente importante: **testa a significância estatística do regressor associado com $\boldsymbol{\beta_j}$**.

O teste é então $\quad H_0: \beta_j = 0$ vs $H_1: \beta_j \neq 0$

Estatística teste: $T = \frac{\hat{\beta}_1 - 0}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t(n-k-1)$

Rejeitar H_0 equivale a aceitar que a variável x_j é relevante para explicar o comportamento de y. Pelo contrário, **não rejeitar $\boldsymbol{H_0}$ equivale a não rejeitar que $\boldsymbol{x_j}$ é irrelevante** para explicar o comportamento de y.

Como este teste é feito de forma rotineira para todos os regressores, o seu resultado figura nos outputs de computador.

Significância de x_j - exemplo

Exemplo: Retome-se o exemplo e teste-se a significância estatística dos regressores.

$$H_0: \beta_j = 0 \text{ vs } H_1: \beta_j \neq 0 \qquad t_{j,obs} = \frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}} \qquad valor - p = P(|T| \geq |t_{j,obs}|)$$

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.794906945					
R Square	0.63187705					
Adjusted R Square	0.623215334					
Standard Error	63.04837628					
Observations	88					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	2	579971.1994	289985.5997	72.95055817	3.58672E-19	
Residual	85	337883.3088	3975.097751			
Total	87	917854.5083				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-19.28550028	31.0475285	-0.621160563	0.536156232	-81.0163048	42.44530424
area(m2)	1.38360615	0.148943494	9.289470184	1.40049E-14	1.08746658	1.67974572
quartos	15.12133684	9.488597692	1.593632413	0.114730383	-3.744537442	33.98721112

Significância de x_j - exemplo

Exemplo (continuação): Mesmo exemplo com output STATA

```
. regress preco area quartos

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  2,    85) =   72.95
       Model |  579971.198     2  289985.599           Prob > F      =  0.0000
    Residual |  337883.308    85  3975.09774           R-squared     =  0.6319
-------------+------------------------------           Adj R-squared =  0.6232
       Total |  917854.506    87  10550.0518           Root MSE      =  63.048

------------------------------------------------------------------------------
       preco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        area |   1.383606   .1489435     9.29   0.000     1.087467    1.679746
     quartos |   15.12134   9.488598     1.59   0.115    -3.744538    33.98721
       _cons |   -19.2855   31.04753    -0.62   0.536    -81.0163     42.4453
------------------------------------------------------------------------------
```

Significância de x_j - exemplo

```
. regress preco quartos area

      Source |       SS       df       MS              Number of obs =      88
-------------+------------------------------           F(  2,    85) =   72.95
       Model |  579971.198     2  289985.599           Prob > F      =  0.0000
    Residual |  337883.308    85  3975.09774           R-squared     =  0.6319
-------------+------------------------------           Adj R-squared =  0.6232
       Total |  917854.506    87  10550.0518           Root MSE      =  63.048

------------------------------------------------------------------------------
       preco |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     quartos |   15.12134   9.488598     1.59   0.115    -3.744538    33.98721
        area |   1.383606   .1489435     9.29   0.000     1.087467    1.679746
       _cons |  -19.2855   31.04753    -0.62   0.536    -81.0163     42.4453
------------------------------------------------------------------------------


. matrix list e(V)

symmetric e(V)[3,3]
            quartos        area       _cons
quartos   90.033486
   area  -.7520268   .02218416
  _cons  -180.54992  -1.4673392   963.94902
```

Teste ao sinal de um β_j

Outro caso particular que surge com alguma frequência é o de avaliar o sinal do impacto. Neste caso o teste será

Caso 1 $- H_0: \beta_j \leq 0 \text{ vs } H_1: \beta_j > 0$ ou $H_0: \beta_j = 0 \text{ vs } H_1: \beta_j > 0$

Caso 2 $- H_0: \beta_j \geq 0 \text{ vs } H_1: \beta_j < 0$ ou $H_0: \beta_j = 0 \text{ vs } H_1: \beta_j < 0$

Observações:

1. Os dois testes não são de todo equivalentes. O caso 1 testa um impacto positivo enquanto o caso 2 testa um impacto negativo (ver H_1). A resposta é "afirmativa" quando se rejeita H_0.
2. Nas 2 situações a estatística de teste será $T = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t(n-k-1)$
3. No 1º caso a região de rejeição é unilateral direita e no segundo caso unilateral esquerda.

Estatística t: principais variantes

Teste à significância (individual) estatística de um regressor:

Hipóteses	t_j	Nota
$H_0: \beta_j = 0$ $H_1: \beta_j \neq 0$	$\dfrac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}}$	Apresentado automaticamente nos output da regressão

Teste ao sinal de um coeficiente:

Hipóteses	t_j	Nota
$H_0: \beta_j = 0$ $H_1: \beta_j > 0$ $(H_1: \beta_j < 0)$	$\dfrac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}}$	Região de rejeição unilateral direita (esquerda)

Teste para um valor particular de um coeficiente (c):

Hipóteses	t_j	Nota
$H_0: \beta_j = c$ $H_1: \beta_j \neq c$ $(H_1: \beta_j > c$ ou $H_1: \beta_j < c)$	$\dfrac{\hat{\beta}_j - c}{\hat{\sigma}_{\hat{\beta}_j}}$	Região de rejeição bilateral (unilateral à direita ou esquerda)

Relevância prática de um regressor

Para além da significância estatística é necessário analisar a relevância prática do regressor, isto é se a variável x_j tem um impacto efetivo no comportamento de y .

Como $t_{j,obs} = \frac{\widehat{\beta}_j}{\widehat{\sigma}_{\widehat{\beta}_j}}$, um valor elevado de $\left|t_{j,obs}\right|$ que leve a uma rejeição clara de $H_0\colon \beta_j = 0$ pode ter origem:

- Num valor elevado de $\left|\beta_j\right|$ (combinado com um valor razoável de $\hat{\sigma}_{\widehat{\beta}_j}$);
- Num valor muito pequeno de $\hat{\sigma}_{\widehat{\beta}_j}$ sem que $\left|\beta_j\right|$ assuma um valor que seja relevante em termos práticos (o que pode ser facilitado por amostras de grande dimensão)

Relevância prática de um regressor

Exemplo Wooldridge – Participation Rates in 401 (k) Plans

A ideia é explicar a taxa de participação *prate* (percentagem de trabalhadores elegíveis que aderem ao plano) no plano de pensões 401 (k) em função do peso da contribuição patronal *mrate* (montante que a empresa mete no fundo por cada dólar de contribuição do trabalhador), da idade do plano *age,* e do tamanho da empresa *totemp* (nº de trabalhadores). Dados na página da UC.

O modelo estimado foi

$$\widehat{prate} = \underset{(0.78)}{80.29} + \underset{(0.52)}{5.44}\, mrate + \underset{(0.045)}{0.269}\, age - \underset{(0.0000367)}{0.0001297}\, totemp$$

$$n = 1534 \quad R^2 = 0.100 \quad \overline{prate} = 87.4 \quad \overline{totemp} = 3567.3$$

Como se pode verificar, rejeita-se claramente $H_0\colon \beta_3 = 0$ $(t_{3,obs} = -3.53)$ logo $totemp$ é estatisticamente significante relevante $(p - value = 0.00042)$ mas de pouca relevância prática: Para um crescimento de 10000 trabalhadores na empresa a tx de participação diminuiria de 1.297pp!

Não olhar tanto para o valor do coeficiente mas sim este valor multiplicado pelos valores “razoáveis” da variável x!

Testes sobre **uma** combinação linear de coeficientes

O propósito é agora fazer inferência (IC ou TH) sobre uma combinação linear de coeficientes $\delta = \beta_0 c_0 + \beta_1 c_1 + \cdots + \beta_k c_k$.

Observações:

1. Os β_j são os coeficientes do modelo que estimámos
2. Os c_j são valores pré-fixados em função do que se pretende testar.
3. Por exemplo, se no modelo

$$preço = \beta_0 + \beta_1 area + \beta_2 quartos + u$$

   Se se quisesse testar se o impacto de 10 m2 de área equivale ao impacto de 1 quarto definir-se-ia $\delta = 10\ \beta_1$- β_2 e testar-se-ia $H_0 \colon \delta = 0$

4. Como é óbvio este caso engloba o anterior (1 parâmetro)

Testes sobre **uma** combinação linear de coeficientes

A estimação de δ não levanta problema

$$\hat{\delta} = \hat{\beta}_0 c_0 + \hat{\beta}_1 c_1 + \cdots + \hat{\beta}_k c_k = c\hat{\beta}$$

Sendo $c = [c_0 \quad c_1 \quad \cdots \quad c_k]$ um vetor linha e $\hat{\beta}$ o vetor coluna dos estimadores/estimativas de β.

Para obter a variância do estimador $\hat{\delta}$, aplica-se a regra habitual para a variância de uma combinação linear de v.a., **não esquecendo as covariâncias** entre os $\hat{\beta}_j$. Em termos matriciais vem

$$var(\hat{\delta}) = \sigma_{\hat{\delta}}^2 = var(c\hat{\beta}) = c\, var(\hat{\beta}) c^T = \sigma^2 c(\boldsymbol{X}^T\boldsymbol{X})^{-1} c^T$$

já que $var(\hat{\beta}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

Assim das propriedades da normal vem $\hat{\delta} \sim N(\delta, \sigma_{\hat{\delta}}^2)$

Testes sobre **uma** combinação linear de coeficientes

Como $\sigma^2_{\hat{\delta}}$ é desconhecido por não se conhecer σ^2, o único elemento desconhecido, aplica-se uma raciocínio em tudo idêntico ao que se aplicou para $\hat{\beta}_j$ e obtém-se

$$t_{\hat{\delta}} = \frac{\hat{\delta}-\delta}{\hat{\sigma}_{\hat{\delta}}} \sim t(n-k-1)$$

com base no qual se fará a inferência estatística.

- IC $(1-\alpha)$ vem $\left(\hat{\delta} - t_{\alpha/2}\hat{\sigma}_{\hat{\delta}};\ \hat{\delta} + t_{\alpha/2}\hat{\sigma}_{\hat{\delta}}\right)$
- TH – Em tudo semelhante ao que se fez para $\hat{\beta}_j$

A dificuldade prática reside em obter $\hat{\sigma}_{\hat{\delta}}$ quer porque muitos softwares não reportam diretamente a matriz estimada das var/cov de $\hat{\beta}$ quer, em menor grau, por obrigar a uma conta adicional.

Testes sobre **uma** combinação linear de coeficientes

Retorne-se ao exemplo

$$\widehat{preço} = -19.286 + 1.3836\, area + 15.121\, quartos$$
$$(31.046) \quad (0.1489) \qquad\qquad (9.4886)$$

e assuma-se que se queria testar se o impacto de 10 m2 de área equivale ao impacto de 1 quarto.

De acordo com o que se viu define-se $\delta = 10\,\beta_1$- β_2 e testa-se $H_0: \delta = 0$ *vs* $H_1: \delta \neq 0$ com base na estatística de teste

$$t = \frac{\hat{\delta}}{\hat{\sigma}_{\hat{\delta}}} \sim t(88 - 2 - 1).$$

Em termos observados, $\hat{\delta} = 10\,\hat{\beta}_1 - \hat{\beta}_2 = -1.285$ e

$$\hat{\sigma}^2_{\hat{\delta}} = 10^2 \hat{\sigma}^2_{\hat{\beta}_1} + \hat{\sigma}^2_{\hat{\beta}_2} - 2 \times 10 \times 1 \hat{\sigma}_{\hat{\beta}_1 \hat{\beta}_2}$$

Para obter $\hat{\sigma}_{\hat{\beta}_1 \hat{\beta}_2}$ tem de recorrer-se à matriz estimada das var/cov de $\hat{\beta}$. Note-se que $\hat{\sigma}^2_{\hat{\beta}_1} = 0.1489^2$ e $\hat{\sigma}^2_{\hat{\beta}_2} = 9.4886^2$.

Testes sobre **uma** combinação linear de coeficientes

Recordando a matriz apresentada anteriormente (ver tópico sobre propriedades do modelo – MLR 5)

	(Intercept)	area	quartos
(Intercept)	963.949026	-1.46733923	-180.5499241
area	-1.467339	0.02218416	-0.7520268
quartos	-180.549924	-0.75202680	90.0334862

$$\hat{\sigma}^2_{\hat{\delta}} = 10^2\hat{\sigma}^2_{\hat{\beta}_1} + \hat{\sigma}^2_{\hat{\beta}_2} - 2 \times 10 \times 1\hat{\sigma}_{\hat{\beta}_1\hat{\beta}_2}$$
$$= 100 \times 0.02218 + 90.033 + 20 \times 0.75203 = 107.29$$

Assim, $t_{\hat{\delta},obs} = \frac{\hat{\delta}}{\hat{\sigma}_{\hat{\delta}}} = \frac{-1.285}{\sqrt{107.29}} = -0.124$ e portanto não se rejeita H_0 ,

Valor-p=0.9015

Vamos utilizar este exemplo para introduzir uma metodologia alternativa para efetuar o mesmo teste

Observação:

Neste caso $\delta = 10\ \beta_1 - \beta_2 = c\beta$ com $c = [0 \quad 10 \quad -1]$ e

$$\hat{\sigma}^2_{\hat{\delta}} = [0 \quad 10 \quad -1] \begin{bmatrix} 963.949 & -1.46734 & -180.550 \\ -1.4674 & 0.022184 & -0.75203 \\ -180.550 & -0.75203 & 90.0334 \end{bmatrix} \begin{bmatrix} 0 \\ 10 \\ -1 \end{bmatrix}$$

$$= [-14.674 + 180.550 \quad 0.22184 + 0.75203 \quad -7.5203 - 90.0334] \begin{bmatrix} 0 \\ 10 \\ -1 \end{bmatrix}$$

$$= 0 + 10 \times (0.22184 + 0.75203) + 1 \times (7.520 + 90.0334)$$

$$= 9.7387 + 97.5534 = 107.2921$$

$$\hat{\sigma}_{\hat{\delta}} = 10.358$$

Testes sobre uma combinação linear de coeficientes

Método alternativo

Temos $\delta = 10\beta_1 - \beta_2$

Resolver em ordem a um dos $\beta_j \rightarrow \beta_2 = 10\beta_1 - \delta$

Obter a regressão auxiliar: Substituir o β_j escolhido no modelo pela expressão obtida e simplificar

$$preço = \beta_0 + \beta_1 area + \beta_2 quartos + u$$

$$preço = \beta_0 + \beta_1 area + (10\beta_1 - \delta\,)\,quartos + u$$

$$preço = \beta_0 + \beta_1 area + 10\beta_1 quartos - \delta\, quartos + u$$

$$preço = \beta_0 + \beta_1 (area + 10quartos) - \delta\, quartos + u$$

$preço = \beta_0 + \beta_1 x^* + \delta\,(-quartos) + u$ com $x^* = (area + 10quartos)$

Definir a nova variável e estimar o modelo auxiliar (slide seguinte)

Fazer inferência sobre δ a partir do output obtido para a regressão auxiliar

Output regressão auxiliar

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.794906945					
R Square	0.63187705					
Adjusted R Square	0.623215334					
Standard Error	63.04837628					
Observations	88					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	2	579971.1994	289985.5997	72.95055817	3.58672E-19	
Residual	85	337883.3088	3975.097751			
Total	87	917854.5083				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-19.28550028	31.0475285	-0.621160563	0.536156232	-81.0163048	42.44530424
x*	1.38360615	0.148943494	9.289470184	1.40049E-14	1.08746658	1.67974572
quartos	-1.285275338	10.35820635	-0.124082809	0.901542671	-21.8801646	19.30961392

Repare-se que a linha de x^* referente ao parâmetro $\hat{\beta}_1$ não se alterou em relação ao modelo inicial (*área*)

Testes sobre uma combinação linear de coeficientes

Exemplo: Considere o modelo (ver o tópico relevância prática de um regressor)

$$prate = \beta_0 + \beta_1 mrate + \beta_2 age + \beta_3 totemp + u$$

a) Escreva a regressão auxiliar para fazer inferência sobre $\delta = 4\beta_1 - 10\beta_2$

b) Mesma questão para $\theta = \frac{\beta_2}{2} + 1000\beta_3$

c) Obter um IC a 95% para δ

d) Testar $H_0: \delta = 0$ contra $H_1: \delta \neq 0$

Testes sobre uma combinação linear de coeficientes

$$prate = \beta_0 + \beta_1 mrate + \beta_2 age + \beta_3 totemp + u$$

a) Escreva a regressão auxiliar para fazer inferência sobre $\delta = 4\beta_1 - 10\beta_2$

Sol 1:

$\delta = 4\beta_1 - 10\beta_2$ logo $\beta_1 = 0.25\,\delta + 2.5\,\beta_2$

$prate = \beta_0 + (0.25\,\delta + 2.5\,\beta_2)mrate + \beta_2 age + \beta_3 totemp + u$

$prate = \beta_0 + \delta\,(0.25 \times mrate) + \beta_2(age + 2.5\,mrate) + \beta_3 totemp + u$

Sol 2:

$\delta = 4\beta_1 - 10\beta_2$ logo $\beta_2 = 0.4\,\beta_1 - 0.1\,\delta$

$prate = \beta_0 + \beta_1 mrate + (0.4\,\beta_1 - 0.1\,\delta\,)age + \beta_3 totemp + u$

$prate = \beta_0 + \beta_1\,(mrate + 0.4\,age) + \delta(-0.1 \times age) + \beta_3 totemp + u$

Desafio: Estime as 2 regressões auxiliares e verifique que as linhas referentes a δ são idênticas.

Testes sobre uma combinação linear de coeficientes

$$prate = \beta_0 + \beta_1 mrate + \beta_2 age + \beta_3 totemp + u$$

b) Mesma questão para $\theta = \frac{\beta_2}{2} + 1000\beta_3$

Sol 1:

$\theta = \frac{\beta_2}{2} + 1000\beta_3$ logo $\beta_2 = 2\,\theta - 2000\,\beta_3$

$prate = \beta_0 + \beta_1 mrate + (2\,\theta - 2000\,\beta_3)age + \beta_3 totemp + u$

$prate = \beta_0 + \beta_1 mrate + \theta\,(2 \times age) + \beta_3(totemp - 2000 \times age) + u$

Sol 2:

$\theta = \frac{\beta_2}{2} + 1000\beta_3$ logo $\beta_3 = 0.001\,\theta - 0.0005\,\beta_2$

$prate = \beta_0 + \beta_1 mrate + \beta_2\,age + (0.001\,\theta - 0.0005\,\beta_2)\,totemp + u$

$prate = \beta_0 + \beta_1 mrate + \beta_2\,(age - 0.0005\,totemp) + \theta\,(0.001 \times totemp) + u$

Testes sobre uma combinação linear de coeficientes

$$prate = \beta_0 + \beta_1 mrate + \beta_2 age + \beta_3 totemp + u$$

c) Obter um IC a 95% para δ - sol 1

$$prate = \beta_0 + \delta\,(0.25 \times mrate) + \beta_2(age + 2.5\,mrate) + \beta_3 totemp + u$$

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.315592361					
R Square	0.099598538					
Adjusted R Square	0.097833045					
Standard Error	15.87778005					
Observations	1534					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	3	42666.57348	14222.19116	56.41400708	1.37991E-34	
Residual	1530	385718.966	252.1038993			
Total	1533	428385.5394				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	80.29428539	0.777695208	103.2464706	0	78.76882403	81.81974675
0.25mrate	19.07165742	2.199785126	8.669781966	1.08405E-17	14.75674437	23.38657047
age+2.5mrate	0.269407284	0.04514857	5.967127681	2.99497E-09	0.180847655	0.357966913
totemp	-0.000129781	3.67184E-05	-3.534503198	0.000420703	-0.000201805	-5.77576E-05

d) $H_0: \delta = 0$ contra $H_1: \delta \neq 0$ $\quad t_{obs} = 8.67$ $\quad p - value \approx 0$ Rejeita-se H_0

Testes sobre q combinações lineares de coeficientes

Vamos agora considerar que queremos testar várias combinações lineares de β_j

Embora se possa continuar a considerar as 2 vias de abordagem anteriores (“matricial” ou regressão auxiliar por reparametrização da regressão original) apenas se irá desenvolver a segunda por ser bastante mais eficiente.

Apresentaremos o problema do caso mais simples para o mais geral (complicado) e utilizaremos um novo exemplo com base numa regressão envolvendo mais parâmetros

Testes sobre *q* combinações lineares de coeficientes

Exemplo: Vamos modelar o logaritmo do salário dos jogadores profissionais de baseball (dados de 1993 no ficheiro MLB1_simplificado.xlsx) em função de 5 características:

- $years$ – nº de anos de carreira do jogador
- $gamesyr$ – nº médio de jogos que o jogador joga por ano
- $bavg$ – nº médio de batting/ano
- $hrunsyr$ – nº médio de *home runs* por ano
- $rbisyr$ – nº médio de “runs” por ano (“runs batted in per year”)

Modelo

$$\ln salary = \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 hrunsyr + \beta_5 rbisyr + u$$

Testes sobre *r* combinações lineares de coeficientes

Modelo estimado

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.792340118					
R Square	0.627802862					
Adjusted R Square	0.622439791					
Standard Error	0.726577259					
Observations	353					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	5	308.9892282	61.79784565	117.0603271	2.93802E-72	
Residual	347	183.1863363	0.527914514			
Total	352	492.1755645				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	11.19242075	0.288822864	38.75185162	4.1871E-128	10.62435701	11.76048449
years	0.068862623	0.012114544	5.684293499	2.7876E-08	0.045035448	0.092689799
gamesyr	0.012552112	0.002646763	4.742438528	3.08865E-06	0.007346394	0.017757829
bavg	0.000978594	0.001103509	0.886802204	0.375799595	-0.001191814	0.003149002
hrunsyr	0.014429519	0.01605698	0.898644642	0.369465108	-0.017151734	0.046010772
rbisyr	0.01076574	0.007174962	1.50045958	0.134404707	-0.003346147	0.024877626

Olhando para o output vê-se que os 3 últimos coeficientes não são, individualmente considerados, estatisticamente significativos. Será que o mesmo se pode afirmar para o conjunto?

Teste à **significância conjunta de *q* regressores**

Considere-se, sem perda de generalidade, que se estava a testar a significância conjunta dos últimos *r* coeficientes. Definindo $p = k - q$, a hipótese em teste escreve-se:

$$H_0\colon \beta_{p+1} = \cdots = \beta_k = 0$$

Estratégia de teste

- Estimar o modelo incorporando a restrição definida em H_0, isto é, estimar o modelo sem considerar as variáveis $x_{p+1}, x_{p+2}, \dots, x_k$.
- Comparar a qualidade do modelo com restrição com a qualidade do modelo inicial (sem a restrição):
  - Se a qualidade for semelhante, pode concluir-se que as *r* variáveis associadas com os coeficientes em teste não acrescentam nada de significativo ao modelo e poderão ser consideradas **conjuntamente não significantes** e, eventualmente, eliminadas.
  - Se a introdução da restrição originar uma quebra de qualidade significativa, a conclusão terá de ser a oposta. As variáveis $x_{p+1}, x_{p+2}, \dots, x_k$ são conjuntamente significantes.
- Problemas em aberto:
  - Como medir a qualidade? Utilizar a soma dos quadrados dos resíduos de cada modelo
  - Como definir a fronteira? Basearmo-nos numa estatística de teste.

Teste à **significância conjunta de *q* regressores**

$H_0: \beta_{p+1} = \cdots = \beta_k = 0$ vs $H_1: \exists \beta_j \neq 0, j = p+1, \dots k$

Ou, mais simplesmente, $H_1: H_0$ falsa

Estimar o modelo inicial com k variáveis explicativas → SSR

Estimar o modelo com q restrições ($k - q$ var expli) → SSR_* (q restrições)

Mostra-se que

$$F = \frac{(SSR_* - SSR)/q}{SSR/(n-k-1)} \sim F(q, n-k-1)$$

Por razões óbvias, a região de rejeição situar-se-á na cauda direita da distribuição

Teste à significância conjunta de q regressores

Como neste caso a variável y é a mesma nos 2 modelos, pode mostrar-se que $\frac{(SSR_*-SSR)/q}{SSR/(n-k-1)} = \frac{(R^2-R_*^2)/q}{(1-R^2)/(n-k-1)}$ e portanto, de forma equivalente, tem-se

$$F = \frac{(R^2 - R_*^2)/q}{(1-R^2)/(n-k-1)} \sim F(q, n-k-1)$$

Sendo o “*” referente ao modelo com restrições.

Para mostrar a igualdade fazer $R^2 = 1 - SSR/SST$ sendo SST o mesmo nos 2 modelos (a variação total apenas depende de y).

$R^2 = 1 - SSR/SST$ logo $SSR = (1-R^2)\ SST$

$$\frac{(SSR_* - SSR)}{SSR} = \frac{((1-R_*^2)SST - (1-R^2)SST)}{(1-R^2)SST} = \frac{(1-R_*^2) - (1-R^2)}{(1-R^2)} = \frac{R^2 - R_*^2}{(1-R^2)}$$

Teste à **significância conjunta de *q* regressores**

Exemplo: $H_0: \beta_3 = \beta_4 = \beta_5 = 0 \ vs \ H_1: \ H_0$ falsa (pelo menos um dos $\beta \neq 0$)

Modelo com a restrição

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.772704072					
R Square	0.597071582					
Adjusted R Square	0.594769134					
Standard Error	0.752731258					
Observations	353					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	2	293.8640431	146.9320216	259.3203218	8.2202E-70	
Residual	350	198.3115214	0.566604347			
Total	352	492.1755645				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	11.22380399	0.108312013	103.6247386	9.966E-265	11.01077971	11.43682826
years	0.071317962	0.012505011	5.703150453	2.50422E-08	0.046723543	0.095912381
gamesyr	0.02017448	0.00134287	15.02340897	1.01913E-39	0.017533371	0.022815589

$$F_{obs} = \frac{(198.311 - 183.186)/3}{183.186/(353-1-5)} = 9.55 \qquad p - value = P\left(F_{3;347} \geq 9.55\right) \approx 0 \quad (4.47E - 06)$$

Ou $F_{obs} = \frac{(0.627803 - 0.597072)/3}{(1-0.627803)/(353-1-5)} = 9.55$ $\qquad F_{0.05} = 2.6306$

Rejeita-se H_0 logo as avariáveis são conjuntamente significantes -> Multicolinearidade

Teste à significância conjunta de *todos* os regressores

Hipóteses em teste

$H_0: \beta_1 = \cdots = \beta_k = 0 \ vs \ H_1: \ H_0$ falsa (pelo menos um dos $\beta \neq 0$)

Utiliza-se a estratégia de teste vista para o caso anterior.

O modelo com restrições não precisa de ser estimado já que não incorporando nenhum regressor teremos: $R_*^2 = 0, \ SSR_* = SST$ e SSE_*=0.

A estatística de teste será então $F = \frac{(SST-SSR)/k}{SSR/(n-k-1)} \sim F(k, n-k-1)$ ou $F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{R^2}{1-R^2}\frac{n-k-1}{k} \sim F(k, n-k-1)$

Como este teste (teste à significância global da regressão) é feito de forma rotineira ela está incluído em todos os output de computador.

Testes sobre várias combinações lineares – caso geral

A metodologia de teste que se apresentou pode ser estendida para qualquer sistema de restrições lineares.

Para tal:

- Definir o sistema de restrições que se quer testar
- Resolver o sistema em ordem a um sub-conjunto de β_j
- Definir a regressão auxiliar e estimá-la
- Calcular F_{obs} e realizar o teste (utilizando o *valor-p* ou valor crítico)

Nota importante: Caso a variável dependente da regressão auxiliar seja diferente da variável dependente da regressão original não se pode utilizar o resultado baseado no R^2.

Testes sobre várias combinações lineares – caso geral

Exemplo: Considere-se o exemplo anterior e admita-se que se quer testar

$$H_0: \begin{cases} \beta_1 = 0.1 \\ \beta_4 - \beta_5 = 0 \end{cases} \text{ vs } H_1: \ H_0 \text{ falsa}$$

- Resolver o sistema

$$\begin{cases} \beta_1 = 0.1 \\ \beta_4 - \beta_5 = 0 \end{cases} \Leftrightarrow \begin{cases} \beta_1 = 0.1 \\ \beta_4 = \beta_5 \end{cases}$$

- Regressão auxiliar

$$\ln salary = \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 hrunsyr + \beta_5 rbisyr + u$$

$$= \beta_0 + 0.1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_5 hrunsyr + \beta_5 rbisyr + u$$

$$\ln salary - 0.1\ years = \beta_0 + \beta_2 gamesyr + \beta_3 bavg + \beta_5 (hrunsyr + rbisyr) + u$$

(ver output no slide seguinte)

- $F_{obs} = \frac{(186.674 - 183.186)/2}{183.186/(353-5-1)} = 3.3037 \qquad F_{0.05}(2;347) = 3.022$

$p - value = P(F \geq 3.3037) = 0.0379$ Rejeita H_0 para $\alpha = 0.05$

Testes sobre várias combinações lineares – caso geral

Exemplo: output da regressão auxiliar

$$\ln salary - 0.1\ years = \beta_0 + \beta_2 gamesyr + \beta_3 bavg + \beta_5(hrunsyr + rbisyr) + u$$

SUMMARY OUTPUT		ln(salary)-0.1*years				
Regression Statistics						
Multiple R	0.691355747					
R Square	0.477972769					
Adjusted R Square	0.473485429					
Standard Error	0.731357498					
Observations	353					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	3	170.9207773	56.97359242	106.5158331	5.65962E-49	
Residual	349	186.6744424	0.534883789			
Total	352	357.5952197				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	11.19091343	0.272181054	41.11569581	7.6181E-136	10.65559194	11.72623493
gamesyr	0.010532168	0.001858351	5.667481593	3.03747E-08	0.006877193	0.014187143
bavg	0.000878062	0.001074004	0.817559631	0.414166471	-0.001234272	0.002990396
hrunsyr+rbisyr	0.011681257	0.002298988	5.081043186	6.12886E-07	0.007159643	0.016202871

Anexos à parte 1

Teorema de Frisch-Waugh - demonstração

Demonstração:

Iremos provar, sem perda de generalidade, para $j = 1$ que $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^{\ 2}}$

Considerem-se as 3 regressões:

1. $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$
2. $\hat{x}_1 = \hat{\alpha}_0 + \hat{\alpha}_2 x_2 + \cdots + \hat{\alpha}_k x_k$ sendo $\hat{r}_{i1}$ os resíduos deste modelo, isto é, $x_{i1} = \hat{x}_{i1} + \hat{r}_{i1}$
3. $\hat{y} = \hat{\gamma}_0 + \hat{\gamma}_1 \hat{r}_{i1}$

Quer-se provar que $\hat{\gamma}_1 = \hat{\beta}_1$.

Parte-se das condições de 1ª ordem na determinação de $\hat{\beta}$ (modelo 1), mais particularmente da equação com $j = 1$

$$\sum_{i=1}^{n} x_{i1}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \cdots - \hat{\beta}_k x_k\right) = \sum_{i=1}^{n} x_{i1}\,(y_i - \hat{y}_i) = \sum_{i=1}^{n} x_{i1}\,\hat{u}_i = 0$$

Substuindo x_{i1} por $\hat{x}_{i1} + \hat{r}_{i1}$ (modelo 2) vem

$$\sum_{i=1}^{n}(\hat{x}_{i1} + \hat{r}_{i1})\,\hat{u}_i = \sum_{i=1}^{n}\hat{x}_{i1}\hat{u}_i + \sum_{i=1}^{n}\hat{r}_{i1}\hat{u}_i = 0 \qquad \text{(A)}$$

Teorema de Frisch-Waugh - demonstração

Mas

$\sum_{i=1}^{n} \hat{x}_{i1}\hat{u}_i = \sum_{i=1}^{n}(\hat{\alpha}_0 + \hat{\alpha}_2 x_{i2} + \cdots + \hat{\alpha}_k x_{ik})\hat{u}_i$

$= \hat{\alpha}_0 \sum \hat{u}_i + \hat{\alpha}_2 \sum x_{i2}\hat{u}_i + \cdots + \hat{\alpha}_k \sum x_{ik}\hat{u}_i = 0$

(propriedades 1 e 2 dos resíduos OLS no mod 1)

$\sum_{i=1}^{n} \hat{r}_{i1}\hat{u}_i = \sum_{i=1}^{n} \hat{r}_{i1}(y_i - \hat{y}_i) = \sum \hat{r}_{i1} y_i - \sum \hat{r}_{i1}\hat{y}_i$

$= \sum \hat{r}_{i1} y_i - \sum \hat{r}_{i1} \left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}\right)$ modelo 1

$= \sum \hat{r}_{i1} y_i - \hat{\beta}_0 \sum \hat{r}_{i1} - \hat{\beta}_1 \sum x_{i1}\hat{r}_{i1} - \hat{\beta}_2 \sum x_{i2}\hat{r}_{i1} - \cdots - \hat{\beta}_k \sum x_{ik}\hat{r}_{i1}$

$= \sum \hat{r}_{i1} y_i - \hat{\beta}_1 \sum x_{i1}\hat{r}_{i1}$

já que $\sum \hat{r}_{i1} = 0$ e $\sum x_{ij}\hat{r}_{i1} = 0$ para $j = 2, \ldots, k$ prop resid modelo 2

$= \sum \hat{r}_{i1} y_i - \hat{\beta}_1 \sum \hat{r}_{i1}^2$ já que $\sum x_{i1}\hat{r}_{i1} = \sum(\hat{x}_{i1} + \hat{r}_{i1})\hat{r}_{i1} = \sum \hat{x}_{i1}\hat{r}_{i1} + \sum \hat{r}_{i1}^2$

e $\sum \hat{x}_{i1}\hat{r}_{i1} = 0$ (prop resid modelo 2)

E portanto, voltando a (A) temos

$\sum \hat{r}_{i1} y_i - \hat{\beta}_1 \sum \hat{r}_{i1}^2 = 0$ isto é $\hat{\beta}_1 = \frac{\sum \hat{r}_{i1} y_i}{\sum \hat{r}_{i1}^2}$ que é a expressão de $\hat{\gamma}_1$ (slide 17 estimação dos coeficientes)

Efeito da adição de uma variável - dem

Demonstração

Considere-se, sem perda de generalidade, o caso em que $j = 1$ e $r = k$. Quer provar-se $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_k \, \tilde{\delta}_1$.

Considerem-se as 2 regressões:

1. $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_{k-1} \, x_{k-1} + \hat{\beta}_k x_k$ modelo completo
2. $y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \cdots + \tilde{\beta}_{k-1} \, x_{k-1} + \tilde{v}$ modelo sem x_k

Aplica-se o Teorema de Frisch-Waugh ao modelo sem x_k , isto é, definindo $\tilde{r}_{i1}$ como sendo os resíduos do modelo $\tilde{x}_1 = \tilde{\alpha}_0 + \tilde{\alpha}_2 x_2 + \cdots + \tilde{\alpha}_{k-1} x_{k-1}$ (3) vem $\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} \tilde{r}_{i1} \, y_i}{\sum_{i=1}^{n} \tilde{r}_{1i}^2}$

Efeito da adição de uma variável - dem

Ora

$$\sum \tilde{r}_{i1}\, y_i = \sum \tilde{r}_{i1}\big(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik} + \hat{u}_i\big)$$ utilizar regressão 1

$$= \hat{\beta}_0 \sum \tilde{r}_{i1} + \hat{\beta}_1 \sum \tilde{r}_{i1} x_{i1} + \cdots + \hat{\beta}_k \sum \tilde{r}_{i1} x_{ik} + \sum \tilde{r}_{i1}\hat{u}_i$$

$$= \hat{\beta}_1 \sum \tilde{r}_{i1} x_{i1} + \hat{\beta}_k \sum \tilde{r}_{i1} x_{ik} + \sum \tilde{r}_{i1}\hat{u}_i$$ já que $\tilde{r}_{i1}$ tem soma nula e é ortogonal a $x_2, x_3, \ldots, x_{k-1}$ (propriedades dos resíduos modelo 3)

$$\sum \tilde{r}_{i1}\, \hat{u}_i = \sum (x_{i1} - \hat{\alpha}_0 - \hat{\alpha}_2 x_{i2} - \cdots - \hat{\alpha}_{k-1} x_{ik-1})\hat{u}_i$$ definição de $\tilde{r}_{i1}$ (mod 3)

$$= \sum x_{i1}\hat{u}_i - \hat{\alpha}_0 \sum \hat{u}_i - \hat{\alpha}_2 \sum x_{i2}\hat{u}_i - \cdots - \hat{\alpha}_{k-1} \sum x_{ik-1}\hat{u}_i$$

$$= 0$$ $\hat{u}_i$ tem soma nula e é ortogonal a $x_2, x_3, \ldots, x_{k-1}$ (prop res mod 1)

$$\sum \tilde{r}_{i1} x_{i1} = \sum \tilde{r}_{i1}\big(\hat{\alpha}_0 + \hat{\alpha}_2 x_{i2} + \cdots + \hat{\alpha}_{k-1} x_{ik-1} + \tilde{r}_{i1}\big)$$ utiliza-se o modelo 3

$$= \hat{\alpha}_0 \sum \tilde{r}_{i1} + \hat{\alpha}_2 \sum \tilde{r}_{i1} x_{i2} + \ldots + \hat{\alpha}_{k-1} \sum \tilde{r}_{i1} x_{ik-1} + \sum \tilde{r}_{i1}^2$$

$$= \sum \tilde{r}_{i1}^2$$ propriedades de $\tilde{r}_{i1}$ vistas anteriormente

E portanto $\tilde{\beta}_1 = \dfrac{\sum_{i=1}^n \tilde{r}_{i1}\, y_i}{\sum_{i=1}^n \tilde{r}_{1i}^2} = \dfrac{\hat{\beta}_1 \sum \tilde{r}_{i1} x_{i1} + \hat{\beta}_k \sum \tilde{r}_{i1} x_{ik} + \sum \tilde{r}_{i1}\hat{u}_i}{\sum \tilde{r}_{i1}^2} = \dfrac{\hat{\beta}_1 \sum \tilde{r}_{i1}^2 + \hat{\beta}_k \sum \tilde{r}_{i1} x_{ik}}{\sum \tilde{r}_{i1}^2}$

Efeito da adição de uma variável - dem

Logo

$$\tilde{\beta}_1 = \frac{\hat{\beta}_1 \sum \tilde{r}_{i1}^2 + \hat{\beta}_k \sum \tilde{r}_{i1} x_{ik}}{\sum \tilde{r}_{i1}^2} = \hat{\beta}_1 + \hat{\beta}_k \left(\frac{\sum \tilde{r}_{i1} x_{ik}}{\sum \tilde{r}_{i1}^2} \right)$$

Como $\tilde{\delta}_1$ é o coeficiente de x_1 na regressão de x_k em $(x_1, x_2, \dots, x_{k-1})$ a regressão auxiliar para aplicar o teorema de Frisch-Waugh será a de x_1 em $(x_2, \dots, x_{k-1})$ que corresponde ao modelo 3 de resíduos $\tilde{r}_{i1}$ e portanto

$$\tilde{\delta}_1 = \frac{\sum_{i=1}^{n} \tilde{r}_{i1}\, x_{ik}}{\sum_{i=1}^{n} \tilde{r}_{1i}^2}\,.$$

Assim $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_k\, \tilde{\delta}_1$ como se queria mostrar